

Непрерывное
обобщение
информационного
критерия Акаике
и его применение
к задаче
выделения
признаков в
линейной
регрессии

Ветров,
Кропотов

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Д. П. Ветров¹ Д. А. Кропотов²

¹МГУ, ВМиК, каф. ММП

²ВЦ РАН

Семинар «Новые подходы в распознавании и
прогнозировании»

Ликбез

Обобщение
критерия Акаике

Выбор модели с
помощью
непрерывного
критерия Акаике

План

Непрерывное
обобщение
информационного
критерия Акаике
и его применение
к задаче
выделения
признаков в
линейной
регрессии

Ветров,
Кропотов

Ликбез

Обобщение
критерия Акаике

Выбор модели с
помощью
непрерывного
критерия Акаике

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Информация по Кульбаку

- Известно, что оптимальные результаты (напр. Байесовский классификатор) в статистическом машинном обучении достигаются при использовании истинной плотности генеральной совокупности
- Один из распространенных подходов при решении задачи МО заключается в оценке условной плотности $p(\mathbf{z})$ по заданной обучающей выборке $Z = \{\mathbf{z}_i\}_{i=1}^n$
- Оценка производится в рамках некоторого параметрического семейства, определяемого параметром θ
- Информация по Кульбаку является мерой качества нашего приближения

$$I(\theta) = \int p(\mathbf{z}) \log p(\mathbf{z}|\theta) d\mathbf{z}$$

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Связь информации по Кульбаку и KL-дивергенции

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

- С точностью до аддитивной константы отрицательная информация по Кульбаку совпадает с дивергенцией Кульбака-Лейблера между $p(\mathbf{z})$ и $p(\mathbf{z}|\boldsymbol{\theta})$ — хорошо известной мерой сходства двух распределений

$$KL = \int p(\mathbf{z}) \log \frac{p(\mathbf{z})}{p(\mathbf{z}|\boldsymbol{\theta})} dz = \int p(\mathbf{z}) \log p(\mathbf{z}) dz - I(\boldsymbol{\theta})$$

- Таким образом, с точки зрения выбора $\boldsymbol{\theta}$ максимизация информации по Кульбаку эквивалентна минимизации дивергенции Кульбака-Лейблера

Альтернативный взгляд

- При использовании метода максимального правдоподобия производится поиск максимума выражения

$$L(\theta) = \log p(Z|\theta) = \sum_{i=1}^n \log p(z_i|\theta), \quad (z) \sim p(z)$$

- Теперь будем увеличивать объем обучающей выборки n . Очевидно, что $n^{-1}L(\theta)$ будет в этом случае все ближе к своему мат. ожиданию

$$\frac{1}{n}L(\theta) \approx \mathbb{E} \log p(z|\theta) = \int p(z) \log p(z|\theta) dz = I(\theta)$$

- Другими словами, информация по Кульбаку есть предел удельного правдоподобия обучающей выборки, если в качестве таковой выступит вся генеральная совокупность

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Информационная матрица Фишера

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез
Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

- Одну из важнейших ролей в математической статистике играет т.н. информационная матрица Фишера F

$$F_{ij} = -\mathbb{E} \frac{\partial^2 \log p(\mathbf{z}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = - \int \frac{\partial^2 \log p(\mathbf{z}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} p(\mathbf{z}) dz$$

- Заметим, что для логарифмически-квадратичной по \mathbf{z} функции правдоподобия, гессииан всюду одинаков, поэтому вместо интегрирования достаточно подсчитать его в одной точке

Распределение оценки максимального правдоподобия

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

- Оценка максимального правдоподобия θ_{ML} зависит от обучающей выборки Z , поэтому является случайной величиной
- При $n \gg 1$ распределение на θ_{ML} близко к нормальному с мат. ожиданием θ_* и дисперсией $(nF)^{-1}$, где

$$\theta_* = \arg \max_{\theta} \int p(x) \log p(x|\theta) dx$$

- Таким образом, оценки максимального правдоподобия являются состоятельными, асимптотически нормальными и асимптотически несмещенными

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Формула Осокина

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

- Для дальнейших выкладок нам понадобится полезная формула для математического ожидания произвольной неотрицательно определенной квадратичной формы по произвольному нормальному распределению
- Пусть $A \geq 0$ и $B \geq 0$ — две неотрицательно определенные квадратные матрицы размера $n \times n$, тогда справедливо

$$\begin{aligned} \sqrt{\frac{\det A}{(2\pi)^n}} \int (\boldsymbol{\theta} - \boldsymbol{\mu}_2)^T B (\boldsymbol{\theta} - \boldsymbol{\mu}_2) \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_1)^T A (\boldsymbol{\theta} - \boldsymbol{\mu}_1)\right) d\boldsymbol{\theta} = \\ = \text{tr}(BA^{-1}) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T B (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \end{aligned}$$

- Заметим, что это формула является обобщением формулы для дисперсии нормального распределения

Формула блочного обращения

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

- При рассмотрении зависимости обратной матрицы от подмножества ее элементов полезна следующая формула

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix}^{-1} = \begin{pmatrix} P^{-1} + P^{-1}QBRP^{-1} & -P^{-1}QB \\ -BRP^{-1} & B \end{pmatrix}$$

- Величина $B = (S - RP^{-1}Q)^{-1}$ называется дополнением Шура (Schur complement)
- Можно показать, что дополнение Шура выражает условную матрицу ковариации в многомерном нормальном распределении, поэтому $B \geq 0$

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез
Элементы теории информации
Свойство оценок максимального правдоподобия
Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Постановка задачи

- Имеется X ресурсов, которые нужно распределить по m целям. Эффект от выделения x_j ресурсов на цель j выражается целевой функцией $h_j(x_j)$
- Как правило предполагается, что целевые функции неубывающие
- Задача максимизировать суммарный эффект

$$\begin{cases} \sum_{j=1}^m h_j(x_j) \rightarrow \max_{x_1, \dots, x_m} \\ \sum_{j=1}^m x_j = X, \quad x_j \leq 0 \end{cases}$$

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез
Элементы теории информации
Свойство оценок максимального правдоподобия
Полезные формулы
Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Лемма Гиббса

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез
Элементы теории информации
Свойство оценок максимального правдоподобия
Полезные формулы
Задача оптимального распределения ресурсов

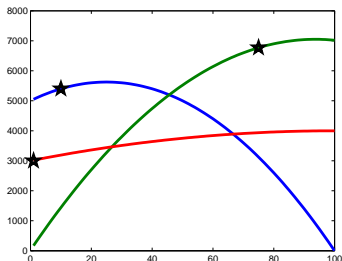
Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Лемма. Пусть \mathbf{x}^0 — оптимальное решение задачи распределения ресурсов и все целевые функции дифференцируемы на участке $[0, X]$. Тогда найдется такое число λ , что выполнено следующее необходимое условие

$$\begin{cases} h_j(x_j^0) = \lambda, & x_j^0 > 0 \\ h_j(x_j^0) \leq \lambda, & x_j^0 = 0 \end{cases}$$

Замечание. Если все целевые функции вогнуты, то это необходимое условие является и достаточным



План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Критерий Акаике

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

- В 1974 году японский математик Акаике предложил свою знаменитую формулу для выбора числа регрессоров в задаче построения линейной регрессии

$$k_* = \arg \max(L(\theta_{ML}) - k),$$

где θ_{ML} — оценка максимального правдоподобия при использовании k регрессоров

- Данная формула явно указывает штраф за каждый дополнительно введенный в модель регрессор, так как понятно, что с увеличением числа регрессоров мы можем объяснить обучающую выборку все точнее и точнее, в том числе и за счет настройки на помехи и шум

Особенности критерия Акаике

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение АИС

Выбор модели с помощью непрерывного критерия Акаике

- Информационный критерий Акаике (AIC) получил огромную популярность в мире по причине своей простоты и одновременной теоретической обоснованности
- Его с успехом применяют не только в линейной регрессии, но и в линейной классификации, кластеризации и др.
- Применение критерия Акаике **необоснованно** в ситуации, когда оцениваемые параметры входят в решающее правило нелинейно, например, в многослойных нейронных сетях
- Критерий Акаике не дает ответа как именно выбрать нужные регрессоры среди множества доступных. Единственным выходом является комбинаторный перебор по всевозможным подмножествам мощности k

Регуляризованный метод максимального правдоподобия

- Для предотвращения перенастройки используется регуляризация

$$\tilde{L}(\boldsymbol{\theta}, \alpha) = \sum_{i=1}^n \log p(\mathbf{z}_i | \boldsymbol{\theta}) - \frac{1}{2} \alpha \|\boldsymbol{\theta}\|^2$$

- Максимизация $\tilde{L}(\boldsymbol{\theta}, \alpha)$ эквивалентна максимизации апостериорной плотности

$$p(\boldsymbol{\theta} | Z, \alpha) = \frac{p(Z | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \alpha)}{\int p(Z | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \alpha) d\boldsymbol{\theta}},$$

где в качестве априорного распределения взято:

$$p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta} | \mathbf{0}, \alpha^{-1})$$

- Значение α определяет количество «эффективных параметров» k в модели — при $\alpha = 0$, k равно числу регрессоров, а при $\alpha = +\infty$, $k = 0$
- Направивается естественное обобщение АИС для задачи выбора наилучшего коэффициента регуляризации α

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение АИС

Выбор модели с помощью непрерывного критерия Акаике

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Оценка обобщающей способности

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

- Пусть у нас имеется обучающая выборка $Z = \{\mathbf{z}_i\}_{i=1}^n$ и тестовая выборка $X = \{\mathbf{x}_i\}_{i=1}^n$
- По обучающей выборке при заданном априорном распределении $p(\boldsymbol{\theta}|A) = \frac{\sqrt{\det(A)}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\boldsymbol{\theta}^T A \boldsymbol{\theta}\right)$ мы находим точку максимума регуляризованного правдоподобия $\boldsymbol{\theta}_{MP}$ и используем ее для распознавания тестовой выборки
- Таким образом правдоподобие тестовой выборки может быть выражено как $\log p(X|\boldsymbol{\theta}_{MP}(Z))$
- Для получения абсолютного критерия достаточно взять мат. ожидания по всевозможным обучающим и тестовым выборкам

$$\mathbb{E}_X \mathbb{E}_Z \log p(X|\boldsymbol{\theta}_{MP}(Z)) \rightarrow \max_A$$

Два предположения о функции правдоподобия

- Предположение 1. Будем считать, что функция правдоподобия логарифмически-квадратична по θ , т.е.

$$\log p(Z|\theta) = \log p(Z|\theta_{ML}) - \frac{1}{2}(\theta - \theta_{ML})^T \nabla^2 \log p(Z|\theta)(\theta - \theta_{ML})$$

Для задачи восстановления линейной регрессии это верно даже для правдоподобия отдельных объектов. При решении задачи классификации предположение разумно при больших n

- Предположение 2. Будем считать, что матрица Фишера может быть адекватно приближена гессианом логарифмического правдоподобия, подсчитанным по выборке длины n

$$F = - \int p(z) \nabla^2 \log p(z|\theta) dz \approx$$

$$-\frac{1}{n} \sum_{i=1}^n \nabla^2 \log p(z_i|\theta) = -\frac{1}{n} \nabla^2 \log p(Z|\theta)$$

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Схема дальнейших выкладок

Непрерывное
обобщение
информационного
критерия Акаике
и его применение
к задаче
выделения
признаков в
линейной
регрессии

Ветров,
Кропотов

Ликбез

Обобщение
критерия Акаике

Критерий
Акаике

Непрерывное
обобщение AIC

Выбор модели с
помощью
непрерывного
критерия Акаике

- Выражение точки θ_{MP} через θ_{ML} , которая не зависит от вида регуляризатора
- Выражение критерия через идеальное значение θ_n^* , которое не зависит от конкретной выборки
- Окончательное получение непрерывного критерия

Шаг 1. Выражение точки θ_{MP} через θ_{ML}

- В критерии фигурирует точка максимума регуляризованного правдоподобия. Чтобы явно выразить ее зависимость от матрицы регуляризации A воспользуемся формулой

$$\theta_{MP} = (H + A)^{-1} H \theta_{ML},$$

где $H = \nabla^2 \log p(Z|\theta)$ — гессиан логарифма правдоподобия выборки

- Точка θ_{MP} является линейной функцией от точки θ_{ML} , поэтому легко определить ее распределение

$$p(\theta_{MP}) = \mathcal{N}(\theta_{MP} | \theta_n^*, (H + A)^{-1} H (H + A)^{-1}),$$

где $\theta_n^* = \arg \max \int p(Z) \log (p(Z|\theta)p(\theta|A)) dZ$ — мат. ожидание точки максимума регуляризованного правдоподобия подсчитанное по всем выборкам объема n

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Шаг 2. Выражение критерия через идеальное значение θ_n^*

Перепишем критерий

$$\begin{aligned} & \int \int p(Z)p(X) \log p(X|\theta_{MP}(Z))dXdZ = \\ & = \int_{\theta_{MP}(Z)=\theta_{MP}} p(\theta_{MP}) \int p(X) \log p(X|\theta_{MP})dXd\theta_{MP} = \\ & = \int p(\theta_{MP}) \int p(X) \log p(X|\theta_{MP})dXd\theta_{MP}. \end{aligned}$$

Разложим внутренний интеграл в ряд Тейлора в окрестности точки θ_n^* и, используя формулу Осокина, получим

$$\begin{aligned} & \int p(X) \log p(X|\theta_n^*)dX + \\ & + \int p(\theta_{MP}) \left[\nabla \int p(X) \log p(X|\theta_n^*)dX \right]^T \times (\theta_{MP} - \theta_n^*)d\theta_{MP} + \\ & + \frac{1}{2} \int p(\theta_{MP}) \left[(\theta_{MP} - \theta_n^*)^T \times \nabla^2 \left(\int p(X) \log p(X|\theta_n^*)dX \right) (\theta_{MP} - \theta_n^*) \right] d\theta_{MP} = \\ & = \int p(X) \log p(X|\theta_n^*)dX - \frac{1}{2} \text{tr}(H(H+A)^{-1}H(H+A)^{-1}). \end{aligned}$$

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Шаг 3. Окончательный вид непрерывного критерия I

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Разложим в ряд Тейлора в окрестности $\theta_{MP}(X)$ функцию $\log p(X|\theta_n^*)$

$$\begin{aligned}\int p(X) \log p(X|\theta_n^*) dX &= \int p(X) \log p(X|\theta_{MP}(X)) dX + \\ &\int p(X) \left[(\nabla \log p(X|\theta_{MP}(X)))^T (\theta_n^* - \theta_{MP}(X)) \right] dX + \\ &\frac{1}{2} \int p(X) \left[(\theta_n^* - \theta_{MP}(X))^T \nabla^2 \log p(X|\theta_{MP}(X)) (\theta_n^* - \theta_{MP}(X)) \right] dX\end{aligned}$$

Учитывая, что $\log p(X|\theta) = -0.5(\theta - \theta_{ML}(X))^T H(\theta - \theta_{ML}(X)) + \text{const}$, можно записать градиент в виде

$$\nabla \log p(X|\theta_{MP}(X)) = -H(\theta_{MP}(X) - \theta_{ML}(X)) = A\theta_{MP}(X)$$

Отсюда

$$\begin{aligned}\int p(X) \left[(\nabla \log p(X|\theta_{MP}(X)))^T (\theta_n^* - \theta_{MP}(X)) \right] dX &= \\ \int_{\theta_{MP}(X)=\theta_{MP}} p(\theta_{MP}) \theta_{MP}^T A (\theta_n^* - \theta_{MP}) d\theta_{MP} &= \{ \theta_n^* = \mathbb{E} \theta_{MP} \} = \\ \int_{\theta_{MP}(X)=\theta_{MP}} p(\theta_{MP}) (\theta_{MP}^T - \theta_n^{*T}) A (\theta_n^* - \theta_{MP}) d\theta_{MP} &= -\text{tr} A(H+A)^{-1} H(H+A)^{-1}\end{aligned}$$

Шаг 3. Окончательный вид непрерывного критерия II

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Применяя формулу Осокина для второго члена разложения в ряд Тэйлора получим

$$\begin{aligned} \frac{1}{2} \int p(X) \left[(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_{MP}(X))^T \nabla^2 \log p(X | \boldsymbol{\theta}_{MP}(X)) (\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_{MP}(X)) \right] dX = \\ \frac{1}{2} \int_{\boldsymbol{\theta}_{MP}(X) = \boldsymbol{\theta}_{MP}} p(\boldsymbol{\theta}_{MP}) \left[(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_{MP})^T \nabla^2 \log p(X | \boldsymbol{\theta}_{MP}) (\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_{MP}) \right] d\boldsymbol{\theta}_{MP} = \\ \frac{1}{2} \text{tr} H(H + A)^{-1} H(H + A)^{-1} \end{aligned}$$

Объединив результаты, полученные на последних двух слайдах окончательно получим

$$\begin{aligned} \mathbb{E}_X \mathbb{E}_Z \log p(X | \boldsymbol{\theta}_{MP}(Z)) = \mathbb{E}_Z \log p(Z | \boldsymbol{\theta}_{MP}(Z)) - \text{tr}(H(H + A)^{-1}) \approx \\ \log p(Z | \boldsymbol{\theta}_{MP}(Z)) - \text{tr}(H(H + A)^{-1}) \end{aligned}$$

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация
CAIC RVM
Общее решение

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Линейная регрессия

- Обобщенная линейная регрессия: по обучающей выборке $(X, \mathbf{t}) = \{(\mathbf{x}_i, t_i)\}_{i=1}^n$, где $\mathbf{x}_i \in \mathbb{R}^d$ — вектор вещественных признаков, $t_i \in \mathbb{R}$ — значение регрессионной переменной, необходимо определить коэффициенты \mathbf{w} линейной формы

$$y = \sum_{j=1}^m w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x})$$

- Максимизация правдоподобия вида

$$p(\mathbf{t}|X, \mathbf{w}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t_i - y_i)^2}{2\sigma^2}\right) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2}\|\mathbf{t} - \Phi\mathbf{w}\|^2\right)$$

эквивалентна использованию метода наименьших квадратов. Здесь $\Phi_{ij} = \phi_j(\mathbf{x}_i)$

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, AIC

β -оптимизация SAIC RVM
Общее решение

Индивидуальные коэффициенты регуляризации

- Будем искать веса линейной регрессии по максимуму апостериорной информации

$$\mathbf{w}_{MP} = \arg \max p(\mathbf{w} | t, X, A) = \arg \max p(t | X, \mathbf{w}) p(\mathbf{w} | A),$$

где

$$p(\mathbf{w} | A) = \frac{\sqrt{\det(A)}}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2} \mathbf{w}^T A \mathbf{w}\right)$$

априорное распределение, играющее роль регуляризатора весов

- Традиционное параметрическое семейство матриц регуляризации $A = \{\alpha I \mid \alpha \geq 0\}$
- При использовании ARD семейство матриц расширяется: $A = \{\text{diag}(\alpha_1, \dots, \alpha_m) \mid \alpha_j \geq 0\}$. Таким образом, каждый вес получает индивидуальный коэффициент регуляризации, который определяет степень адаптируемости конкретного веса под исходные данные

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация SAIC RVM
Общее решение

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Настройка α_j

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Используем непрерывный критерий Акаике (CAIC) для настройки α_j

$$\alpha = \arg \max CAIC(\alpha) = \arg \max (\log p(\mathbf{t}|X, \mathbf{w}_{MP}) - \text{tr}(H(H + A)^{-1}))$$

Прямая оптимизация критерия по α затруднительна, поэтому воспользуемся покоординатным спуском.

Учитывая, что $\mathbf{w}_{MP} = (H + A)^{-1}H\mathbf{w}_{ML}$ можно переписать первое слагаемое

$$\begin{aligned} \log p(\mathbf{t}|X, \mathbf{w}_{MP}) &= \log p(\mathbf{t}|X, \mathbf{w}_{ML}) - \frac{1}{2}(\mathbf{w}_{MP} - \mathbf{w}_{ML})^T H(\mathbf{w}_{MP} - \mathbf{w}_{ML}) = \\ &= \text{const} - \frac{1}{2}\mathbf{w}_{ML}^T ((H + A)^{-1}H - I)H((H + A)^{-1}H - I)\mathbf{w}_{ML} \end{aligned}$$

Теперь α входит только в выражение для $(H + A)^{-1}$

Вспомогательная переменная

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

- Представим матрицу $(H + A)$ в следующем виде

$$(H + A) = \begin{pmatrix} P & \mathbf{q} \\ \mathbf{q}^T & h_{mm} + \alpha_m \end{pmatrix}$$

- Воспользуемся формулой блочного обращения

$$(H + A)^{-1} = \begin{pmatrix} P^{-1} + \beta_m P^{-1} \mathbf{q} \mathbf{q}^T P^{-1} & -\beta_m P^{-1} \mathbf{q} \\ -\beta_m \mathbf{q}^T P^{-1} & \beta_m \end{pmatrix},$$

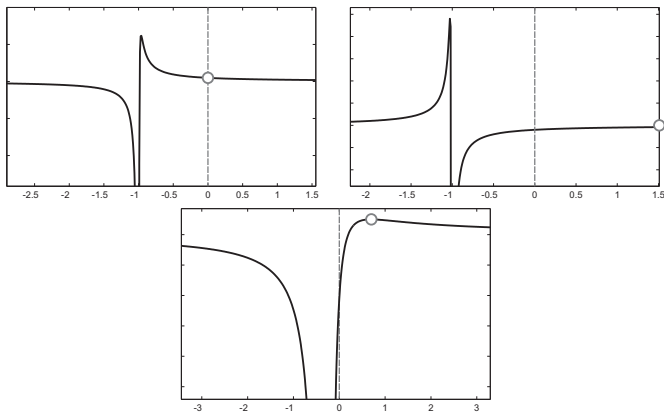
где $\beta_m = (h_{mm} + \alpha_m - \mathbf{q}^T P^{-1} \mathbf{q})^{-1}$ — одномерное дополнение Шура

- Теперь можно выписать зависимость CAIC от β_m . Она оказывается квадратичной

$$CAIC = \text{const} - a^2 \beta_m^2 + b \beta_m$$

Возможные значения α

Легко получить формулу для оптимального значения β_m .
Остается учесть условие, что соответствующее значение α_m
должно быть неотрицательным



Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез
Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация
CAIC RVM
Общее решение

Формулы пересчета

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

- Окончательные формулы для α_m выглядят следующим образом

$$\alpha_m = \begin{cases} \alpha_m^*, & \alpha_m^* \geq 0, \\ 0, & 0 > \alpha_m^* > \mathbf{q}^T \mathbf{P}^{-1} \mathbf{q} - h_{mm}, \\ +\infty, & \alpha_m^* < \mathbf{q}^T \mathbf{P}^{-1} \mathbf{q} - h_{mm}. \end{cases}$$

Здесь α_m^* — значение, соответствующее оптимальному β_m

- Процедура пересчета повторяется итеративно для всех α_j до сходимости

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация
CAIC RVM

Общее решение

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Выделение признаков

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

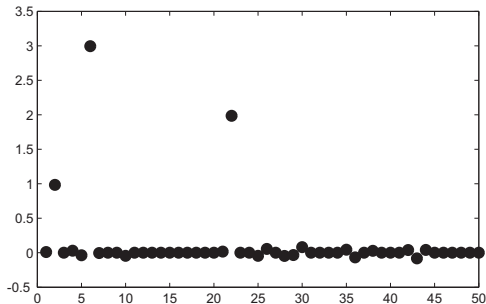
β -оптимизация

CAIC RVM

Общее решение

В выборке 100 объектов, 49 признаков. Данные получены с помощью модели

$$t = x_2 + 3x_6 + 2x_{22} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\varepsilon|0, 0.5), \quad x_i \sim \mathcal{N}(x_i|0, 1)$$



Сравнение с Байесовским ARD

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

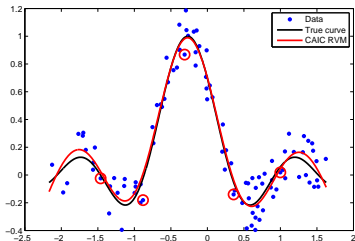
Автоматическое определение значимости, ARD

β -оптимизация CAIC RVM

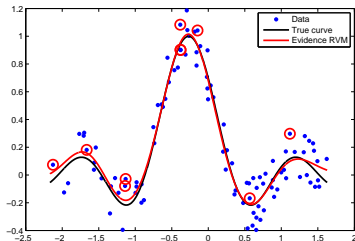
Общее решение

В выборке 100 объектов, сгенерированных с помощью модели:

$$t = \text{sinc}(x) + \varepsilon, \quad \varepsilon \sim \mathcal{R}(-0.2, 0.2), \quad x_i \sim \mathcal{R}(-10, 10)$$



CAIC RVM



EVIDENCE RVM

Разреженность и доля нулевых α_j

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

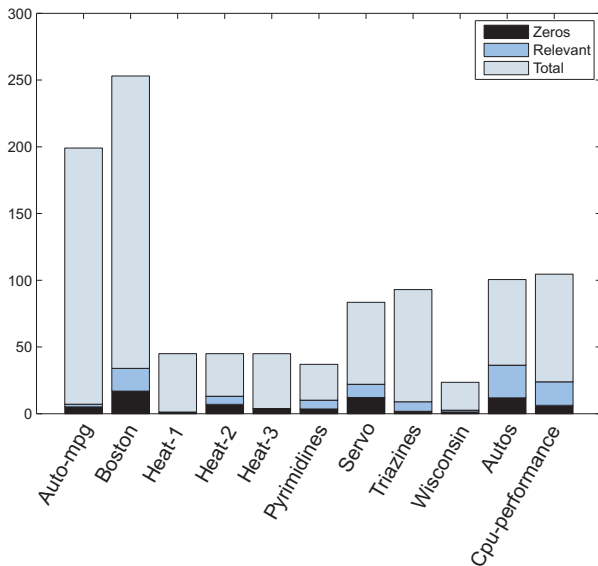
Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение



Выводы

Непрерывное
обобщение
информационного
критерия Акаике
и его применение
к задаче
выделения
признаков в
линейной
регрессии

Ветров,
Кропотов

Ликбез

Обобщение
критерия Акаике

Выбор модели с
помощью
непрерывного
критерия Акаике

Автоматическое
определение
значимости,
ARD

β -оптимизация

CAIC RVM

Общее решение

- Процедура ARD применима не только в Байесовском обучении, но и при использовании других разумных оценок обобщающей способности
- Байесовский и теоретико-информационный подходы являются косвенными проявлениями одного и того же явления
- При теоретико-информационном подходе некоторая часть коэффициентов регуляризации тождественно равна нулю

План

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация CAIC RVM

Общее решение

Ликбез

Элементы теории информации

Свойство оценок максимального правдоподобия

Полезные формулы

Задача оптимального распределения ресурсов

Обобщение критерия Акаике

Критерий Акаике

Непрерывное обобщение AIC

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация

CAIC RVM

Общее решение

Общая постановка задачи и единственность решения

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация SAIC RVM

Общее решение

- Рассмотрим общую постановку задачи, позволив матрице A быть произвольной симметричной неотрицательно определенной

$$\begin{cases} \log p(\mathbf{t}|X, \boldsymbol{\theta}_{MP}) - \text{tr}(H(H + A)^{-1}) \rightarrow \max_A, \\ A = A^T \succeq 0. \end{cases} \quad (1)$$

- В отличие от предыдущего случая, эта задача может быть решена в явном виде
- **Теорема.** Критерий (1) имеет единственный максимум
- **Следствие.** SAIC имеет единственный максимум

Аналитическое решение общей задачи I

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация
CAIC RVM

Общее решение

- Перепишем (1), выразив \mathbf{w}_{MP} через \mathbf{w}_{ML} и матрицу A . С точностью до константы получим

$$-\frac{1}{2}\mathbf{w}_{ML}^T H(H+A)^{-1}H(H+A)^{-1}H\mathbf{w}_{ML} + \mathbf{w}_{ML}^T H(H+A)^{-1}H\mathbf{w}_{ML} - \text{tr}(H(H+A)^{-1}) \rightarrow \max_{A=A^T \succeq 0} \quad (2)$$

- Обозначим $R = H^{1/2}(H+A)^{-1}H^{1/2}$, $\mathbf{v} = H^{1/2}\mathbf{w}_{ML}$. Тогда можно переписать (1) в виде

$$\begin{cases} -\frac{1}{2}\mathbf{v}^T R R \mathbf{v} + \mathbf{v}^T R \mathbf{v} - \text{tr} R \rightarrow \max_R \\ R^{-T} = R^{-1} \succeq I. \end{cases}$$

- Этот функционал вогнут по матрице R

Аналитическое решение общей задачи II

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров, Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация CAIC RVM

Общее решение

- Пусть $R = Q\Lambda Q^T$, $Q^{-1} = Q^T$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$, и $\mathbf{x} = Q\mathbf{v}$. Тогда задачу можно записать в следующем виде

$$\begin{cases} g(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{j=1}^m \left(-\frac{1}{2}x_j^2 \lambda_j^2 + x_j^2 \lambda_j - \lambda_j \right) \rightarrow \max_{\mathbf{x}, \boldsymbol{\lambda}} \\ \sum_{j=1}^m x_j^2 = \|\mathbf{v}\|^2 \\ 0 \leq \lambda_j \leq 1 \quad \forall j = \overline{1, m} \end{cases} \quad (3)$$

- Оптимизируем $g(\mathbf{x}, \boldsymbol{\lambda})$ по $\boldsymbol{\lambda}$ при фиксированном \mathbf{x}

$$\lambda_j^*(\mathbf{x}) = \max \left(0, 1 - \frac{1}{x_j^2} \right). \quad (4)$$

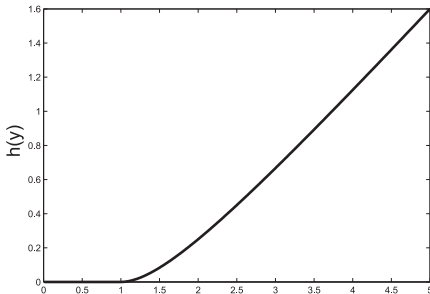
Аналитическое решение общей задачи III

- Пусть $y_j = x_j^2$, введем выпуклую целевую функцию $h(y)$

$$h(y) = \begin{cases} \frac{y}{2} + \frac{1}{2y} - 1, & y \geq 1; \\ 0, & 0 \leq y \leq 1; \end{cases} \quad (5)$$

- После подстановки выражения для $\lambda_j^*(\mathbf{x})$ в задачу оптимизации, получаем стандартную **задачу оптимального распределения ресурсов**, в которой целевая функция $h(y)$ определяет эффективность вложения количества ресурсов y

$$\begin{cases} \sum_{j=1}^m h(y_j) \rightarrow \max_y \\ \sum_{j=1}^m y_j = \|\mathbf{v}\|^2. \end{cases}$$



Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация CAIC RVM

Общее решение

Условие релевантности

- Возможны два случая. Условие $\|\mathbf{v}\| > 1$ явно определяет, содержится ли в обучающей выборке полезная информация с точки зрения выбранного семейства моделей
- При $\|\mathbf{v}\| > 1$ все ресурсы должны быть выделены единственной цели в силу строгой выпуклости функции $h(y)$ при $y > 1$. Не ограничивая общности (все цели имеют одинаковые целевые функции) предположим, что это первая цель, т.е. $y_1 = \|\mathbf{v}\|^2$, $y_2 = \dots = y_m = 0$. Тогда первый собственный вектор матрицы R пропорционален \mathbf{v} , в то время как остальные имеют нулевые собственные значения в силу ограничений (4) и могут быть выбраны произвольным образом, так чтобы $Q^T = Q^{-1}$. Из определения матрицы R получаем искомую матрицу регуляризации

$$A^{-1} = \frac{\mathbf{w}_{ML}^T H \mathbf{w}_{ML} - 1}{\mathbf{w}_{ML}^T H \mathbf{w}_{ML}} \mathbf{w}_{ML} \mathbf{w}_{ML}^T.$$

- При $\|\mathbf{v}\| \leq 1$ все $h(y_j) = 0$, следовательно $\lambda_j^* = 0$, т.е. $A^{-1} = O$, а значит регуляризатор бесконечно большой и все веса решающего правила нулевые $w_j = 0$.

Непрерывное обобщение информационного критерия Акаике и его применение к задаче выделения признаков в линейной регрессии

Ветров,
Кропотов

Ликбез

Обобщение критерия Акаике

Выбор модели с помощью непрерывного критерия Акаике

Автоматическое определение значимости, ARD

β -оптимизация CAIC RVM

Общее решение