Государственный экзамен по специальности

Петр Алексеевич Остроухов

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

18 января 2019 г.

Вопросы

1. Байесовская теория классификации. Линейный дискриминант Фишера. Метод окна Парзена.
2. Задачи геометрического поиска и критерии его эффективности. Алгоритмы локализации точки в простом и выпуклом многоугольниках.
3. Тематическое моделирование. Один из способов вывода EM-алгоритма, регуляризация и модальности, примеры регуляризаторов.

Байесовская теория классификации. Линейный дискриминант Фишера. Метод окна Парзена.

Байесовская теория классификации

X — множество объектов, Y — конечное множество классов, $X \times Y$ — вероятностное пространство с плотностью распределения $p(x, y)$.

- **Дано:** i.i.d. выборка $X^l = (x_i, y_i)_{i=1}^l \sim p(x, y)$;
- **Найти:** алгоритм $a : X \to Y$ с минимальной вероятностью ошибки.

Временное допущение: Пусть известна совместная плотность распределения:

$$p(x, y) = p(x)P(y|x) = P(y)p(x|y).$$

Принцип максимума апостериорной вероятности:

$$a(x) = \arg\max_{y \in Y} P(y|x) = \arg\max_{y \in Y} P(y)p(x|y).$$

Байесовская теория классификации

Теорема (Оптимальный байесовский классификатор)

Пусть $P(y)$ и $p(x|y)$ известны, $\lambda_y \geq 0$ — потеря от ошибки на объекте класса $y \in Y$. Тогда минимум среднего эмпирического риска:

$$R(a) = \sum_{y \in Y} \lambda_y \int [a(x) \neq y] p(x, y) dx$$

достигается байесовским классификатором

$$a(x) = \arg\max_{y \in Y} \lambda_y P(y) p(x|y).$$

Байесовская теория классификации

Подзадачи

1. Восстановление плотности распределения по выборке
   Дано: $X^l = (x_i, y_i)_{i=1}^l$ — обучающая выборка.
   Найти: эмпирические оценки $\hat{P}(y)$ и $\hat{p}(x|y)$, $y \in Y$.
2. Построение классификатора
   Дано: эмпирические оценки $\hat{P}(y)$ и $\hat{p}(x|y)$, $y \in Y$
   (байесовский классификатор здесь уже не оптимален).
   Найти: классификатор $a : X \times Y$, минимизирующий $R(a)$.

Байесовская теория классификации

Наивный байесовский классификатор

Допущение: признаки $f_j : X \to D_j$ — независимы друг от друга, $f_j \sim p_j(\xi|y),\ y \in Y,\ j = 1, ..., n$.
Тогда функции правдоподобия классов представимы в виде:

$$p(x|y) = p_1(\xi_1|y) \cdots p_n(\xi_n|y).$$

В итоге, наш классификатор:

$$a(x) = \arg\max_{y \in Y} \left(\log \lambda_y \hat{P}(y) + \sum_{i=1}^{n} \log \hat{p}(\xi_i|y) \right)$$

Непараметрическое восстановление плотности

По выборке $X^m = (x_i)_{i=1}^m$ оценить плотность $\hat{p}(x)$.
Дискретный случай: $X \gg m$. Гистограмма:

$$\hat{p}(x) = \frac{1}{m} \sum_{i=1}^{m} [x_i = x].$$

Одномерный непрерывный случай: $X = \mathbb{R}$. Эмпирическая оценка по окну ширины h:

$$\hat{p}_h(x) = \frac{1}{mh} \sum_{i=1}^{m} \frac{1}{2} \Big[\frac{|x - x_i|}{h} < 1 \Big].$$

Непараметрическое восстановление плотности

Обобщение: оценка Парзена-Розенблатта по окну h:

$$\hat{p}_h(x; X^m) = \frac{1}{mh} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right),$$

где $K(r)$ — ядро:

- четная;
- нормированная: $\int K(r)dr = 1$;
- невозрастающая при $r > 0$, неотрицательная функция.

Непараметрическое восстановление плотности

Многомерное обобщение: $\rho(x, x')$ — метрика на X. Тогда парзеновская оценка плотности для каждого класса $y \in Y$ в отдельности:

$$\hat{p}_h(x|y) = \frac{1}{l_y V(h)} \sum_{i:y_i=y} K\left(\frac{\rho(x, x_i)}{h}\right)$$

Классификация **методом окна Парзена**:

$$a(x; h) = \arg\max_y \frac{\lambda_y P(y)}{l_y V(h)} \sum_{i:y_i=y} K\left(\frac{\rho(x, x_i)}{h}\right),$$

- $V(h) = \int_X K(\frac{\rho(x,x_i)}{h}) dx$ не должен зависеть от x_i и y_i.
- проблема выбора $K(x)$, h, $\rho(x, x_i)$.

Непараметрическое восстановление плотности

Замечания

- выбор ядра не сильно влияет на качество восстановления плотности при $m \to \infty$;
- выбор h существенно влияет на качество.

Проблема: при наличии локальных сгущений любая h не оптимальна.
Идея: задавать не ширину окна h, а число соседей k, и тогда:

$$h_k(x) = \rho(x, x^{(k+1)}),$$

$x^{(i)}$ — i-й сосед, среди ранжированных по расстоянию от x.
При этом $V(h)$ не должна зависеть от y, поэтому выборка ранжируется целиком, а не по классам.

Параметрическое восстановление плотности

Задана параметрическая модель плотности

$$p(x) = \phi(x;\theta),$$

где θ — параметр, ϕ — фиксированная функция.
Нужно найти оптимальное значение θ по по i.i.d. выборке.
Принцип максимума правдоподобия:

$$L(\theta; X^m) = \sum_{i=1}^{m} \log \phi(x_i;\theta) \to \max_{\theta}.$$

Необходимое условие оптимума:

$$\frac{\partial L(\theta; X^m)}{\partial \theta} = \sum_{i=1}^{m} \frac{\partial \log \phi(x_i;\theta)}{\partial \theta} = 0,$$

где функция $\phi(x;\theta)$ — достаточно гладкая по θ.

Параметрическое восстановление плотности

Пусть $X \subset \mathbb{R}^n$.
Гипотеза: классы имеют n-мерные гауссовские плотности:

$$p(x|y) = \mathcal{N}(x; \mu_y, \Sigma_y) = \frac{\exp(-\frac{1}{2}(x-\mu_y)^T \Sigma_y^{-1}(x-\mu_y))}{\sqrt{(2\pi)^n \det \Sigma_y}}, \;\; y \in Y,$$

где μ_y — вектор матожидания (центр) класса $y \in Y$, Σ_y — симметричная, положительно определенная, невырожденная коварицeонная матрица класса $y \in Y$.

Теорема

1. Разделяющая поверхность $\{x \in X | \lambda_y P(y) p(x|y) = \lambda_s P(s) p(x|s)\}$ квадратична $\forall y, s \in Y, \;\; y \neq s$.
2. Если $\Sigma_y = \Sigma_s$, то она вырождается в линейную.

Параметрическое восстановление плотности:

Теорема

Оценки максимального правдоподобия для n-мерных гауссовских плотностей классов $y \in Y$:

$$\hat{\mu}_y = \frac{1}{l_y} \sum_{i:y_i=y} x_i;$$

$$\hat{\Sigma}_y = \frac{1}{l_y} \sum_{i:y_i=y} (x_i - \hat{\mu}_y)(x_i - \hat{\mu}_y)^T.$$

Квадратичный дискриминант:

$$a(x) = \arg\max_y \log \lambda_y P(y) - \frac{1}{2}(x - \hat{\mu}_y)^T \hat{\Sigma}_y (x - \hat{\mu}_y) - \frac{1}{2} \det \hat{\Sigma}_y.$$

Проблема: для малочисленных классов возможно $\det \hat{\Sigma}_y = 0$.

Параметрическое восстановление плотности:

Допущение: ковариационные матрицы классов равны: $\Sigma_y = \Sigma,\ \forall y \in Y$.

Оценка максимума правдоподобия для Σ:

$$\hat{\Sigma} = \frac{1}{l}\sum_{i=1}^{l}(x_i - \hat{\mu}_{y_i})(x_i - \hat{\mu}_{y_i})^T.$$

Линейный дискриминант Фишера:

$$a(x) = \arg\max_y \lambda_y \hat{P}(y)\hat{p}(x|y) =$$

$$= \arg\max_y(\log(\lambda_y \hat{P}(y)) - \frac{1}{2}\hat{\mu}_y^T \hat{\Sigma}_y^{-1}\hat{\mu}_y + x^T \hat{\Sigma}_y^{-1}\hat{\mu}_y);$$

$$a(x) = \arg\max_y(x^T \alpha_y + \beta_y).$$

Задачи геометрического поиска и критерии его эффективности. Алгоритмы локализации точки в простом и выпуклом многоугольниках.

Задачи геометрического поиска и критерии его эффективности.

Задача поиска

Дано: набор данных, именуемый *файлом* и некоторый новый элемент данных, именуемые *образцом*.
Нужно установить связь между образцом и файлом. Поисковое сообщение, в соответствии с которым ведется просмотр файлы, называется *запросом*.

Особенности геометрического поиска

:

- данные отображают сложные структуры (полигоны, полиэдры и т.п.);
- результатом поиска может оказаться не элемент файла, а положение образца относительно файла.

Задачи геометрического поиска и критерии его эффективности.

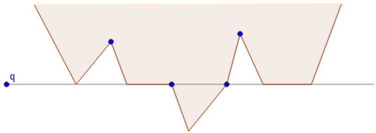
Анализ требуемых ресурсов оценивается по четырем категориям:

1. *Время запроса*. (Сколько времени необходимо на один запрос?)
2. *Память*. (Сколько памяти необходимо для структуры данных?)
3. *Время предобработки*. (Сколько времени потребуется на включение элемента данных в структуру или на удаление из нее?)

Главные модели геометрического поиска:

- *Задача локализации*. Файл — разбиение геометрического пространства, запрос — точка.
- *Задача регионального поиска*. Файл содержит набор точек пространства. Запрос — некоторая стандартная геометрическая фигура, произвольно перемещаемая в этом пространстве.

Локализация точки в простом многоугольнике



Даны простой многоугольник P и точка z. Определить, находится ли q внутри P.

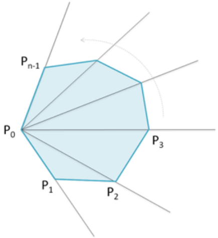
- $L = 0$
- *for i in (1, N) do* (цикл по всем ребрам)
- *if* (ребро[i] не горизонтально) *and* (ребро[i]) пересекает l нижним концом справа от q
  - then $L = L + 1$
- *if* L нечетно *then* q внутри, иначе — снаружи.

Локализация точки в простом многоугольнике

Теорема

Принадлежность точки внутренней области простого n-угольника можно установить за время $O(n)$ без предобработки.

Локализация точки в выпуклом многоугольнике



Предобработка: расположить точки в векторе в направлении по часовой стрелки.
Провести из первой точки лучи через все остальные вершины. Далее необходимо понять, в каком из полученных углов лежит точка. В конце концов, необходимо понять, по какую сторону от ребра сектора лежит данная точка.

Локализация точки в выпуклом многоугольнике

- Предобработка требует времени $O(n)$
- Выполнение запроса требует $O(\log n)$ для поиска сектора, содержащего z, и постоянного времени для локализации точки внутри треугольника сектора.

Теорема

Время ответа на запрос о принадлежности точки выпуклому n-угольнику равно $O(\log n)$ при затратах $O(n)$ памяти и $O(n)$ времени на предобработку.

Тематическое моделирование. Один из способов вывода EM-алгоритма, регуляризация и модальности, примеры регуляризаторов.

Тематическое моделирование (TM)

- W — конечное множество слов
- D — конечное множество документов
- T — конечное множество тем (скрыто)
- $D \times W \times T$ — дискретная генеральная совокупность
- гипотеза условной независимости: $p(w|t,d) = p(w|t)$

Вероятностная тематическая модель порождения текста:

$$p(w|d) = \sum_{t \in T} p(w|t,d)p(t|d) = \sum_{t \in T} p(w|t)p(t|d) = \sum_{t \in T} \phi_{wt}\theta_{td}$$

ТМ [ДНК]

- **Дано**: n_{dw} — частоты слов в документах $\Rightarrow \hat{p}(w|d) = \frac{n_{dw}}{n_d}$
- **Найти**:
  - $\phi_{wt} = p(w|t)$ — вероятности слов в темах
  - $\theta_{td} = p(t|d)$ — вероятности тем в документах
- **Критерий** максимизации правдоподобия:

$$\begin{aligned}\mathcal{L}(\Phi;\Theta) &= \ln p\big((d,w)_{d\in D, w\in W};\Phi,\Theta\big) = \\ &= \ln \prod_{w\in W}\prod_{d\in D}\big(p(w|d){\color{red}p(d)}\big)^{n_{dw}} = \\ &= \sum_{w\in W}\sum_{d\in D} n_{dw}\ln\sum_{t\in T}\phi_{wt}\theta_{td} \to \max_{\Phi;\Theta}\end{aligned}$$

$$\sum_{w\in W}\phi_{wt} = 1,\ \ \phi_{wt} \geq 0,\ \ \sum_{t\in T}\theta_{td} = 1,\ \ \theta_{td} \geq 0.$$

TM [ARTM]

ММП + ARTM:

$$\sum_{d,w} n_{dw} \log \sum_{t \in T} \phi_{wt}\theta_{td} + R(\Phi, \Theta) \to \max_{\Phi,\Theta}; \quad R(\Phi, \Theta) = \sum_{i=1}^{k} \tau_i R_i(\Phi, \Theta)$$

$$\sum_{w \in W} \phi_{wt} = \{0, 1\},\ \phi_{wt} \geq 0,\ \sum_{t \in T} \theta_{td} = \{0, 1\},\ \theta_{td} \geq 0.$$

ТМ [ЕМ-алгоритм]

ЕМ-алгоритм — метод простых итераций для системы уравнений

$$\sum_{d,w} n_{dw} \log \sum_{t\in T} \phi_{wt}\theta_{td} + R(\Phi,\Theta) \to \max_{\Phi,\Theta}; \quad R(\Phi,\Theta) = \sum_{i=1}^{k} \tau_i R_i(\Phi,\Theta)$$

$$\sum_{w\in W} \phi_{wt} = \{0,1\},\ \phi_{wt} \geq 0,\ \sum_{t\in T} \theta_{td} = \{0,1\},\ \theta_{td} \geq 0.$$

Е-шаг:

$$p_{tdw} = \mathrm{norm}_{t\in T}(\phi_{wt}\theta_{td})$$

М-шаг:

$$\phi_{wt} = \mathrm{norm}_{t\in T}\left(n_{wt} + \phi_{wt}\frac{\partial R}{\partial \phi_{wt}}\right), \quad n_{wt} = \sum_{d\in D} n_{dw} p_{tdw};$$

$$\theta_{td} = \mathrm{norm}_{t\in T}\left(n_{td} + \theta_{td}\frac{\partial R}{\partial \theta_{td}}\right), \quad n_{td} = \sum_{w\in d} n_{dw} p_{tdw}.$$

Решение может быть вырожденным для некоторых тем (столбцов матриц Φ) и документов (столбцов матрицы Θ).

Тема t вырождена, если для всех термов $w \in W$

$$n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt} \leq 0}$$

Если тема вырождена, то $\phi_{wt} \equiv 0$; т.е. происходит отбор тем.

Документ d вырожден, если для всех тем $t \in T$

$$n_{td} + \theta_{td} \frac{\partial R}{\partial \theta_{td}} \leq 0$$

Если документ вырожден, то $\theta_{td} \equiv 0$; что означает, что модель не может описать данный документ.

Условия ККН

$$\begin{cases} f(x) \to \min\limits_{x}; \\ g_i(x) \leq 0, \quad i = 1, \ldots, m; \\ h_j(x) = 0, \quad j = 1, \ldots, k. \end{cases}$$

Необходимые условия. Если x — точка локального минимума, то существуют множители μ_i, $i = 1, \ldots, m$, $\lambda_j, j = 1, \ldots, k$:

$$\frac{\partial \mathscr{L}}{\partial x} = 0, \quad \mathscr{L}(x; \mu, \lambda) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{k} \lambda_j h_j(x);$$

$$g_i(x) \leq 0; h_j(x) = 0; \text{ (исходные ограничения)}$$

$$\mu_i \geq 0; \text{ (двойственные ограничения)}$$

$$\mu_i g_i(x) = 0; \text{ условия дополняющей нежесткости}$$

ТМ [ЕМ-алгоритм]

1. Условия ККТ для ϕ_{wt} (аналогично для $\theta_t d$):
$\sum_d n_{dw} \frac{\theta_{td}}{p(w|d)} + \frac{\partial R}{\partial \phi_{wt}} = \lambda_t - \mu_{wt}; \quad \mu_{wt} \geq 0; \quad \mu_{wt}\phi_{wt} = 0.$
2. Умножим обе части равенства на ϕ_{wt} и выделим p_{tdw}:
$\phi_{wt}\lambda_t = \sum_d n_{dw} \frac{\phi_{wt}\theta_{td}}{p(w|d)} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}} = n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}}.$
3. Если $\lambda_t \leq 0$, то тема t вырождена, $\phi_{wt} \equiv 0$ для всех w.
4. Если $\lambda_t > 0$, то либо $\phi_{wt} = 0$, либо $n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}} > 0$:
$$\phi_{wt}\lambda_t = \left(n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}} \right)_+.$$
5. Суммируем обе части равенства по $w \in W$:
$$\lambda_t = \sum_{w \in W} \left(n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}} \right)_+$$
6. Подставив λ_t из (5) в (4), получим результат.

ТМ [модальности]

Пусть M — множество модальностей. Тогда
$\forall m \in M \; \exists W_m : \; \forall i \neq j \; W_i \bigcap W_j = \emptyset, \; \bigcup_{m \in M} W_m = W$.
Тематическая модель для каждой модальности остается без изменений.

ММП для мультимодальной модели:

$$\sum_{m \in M} \tau_m \sum_{d \in D} \sum_{w \in W_m} n_{dw} \ln \sum_{t \in T} \phi_{wt} \theta_{td} \to \max_{\Phi, \Theta},$$

$$\sum_{w \in W} \phi_{wt} = \{0, 1\}, \; \phi_{wt} \geq 0, \; \sum_{t \in T} \theta_{td} = \{0, 1\}, \; \theta_{td} \geq 0,$$

где τ_m — веса модальностей.