МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ ПРОГНОЗИРОВАНИЯ

Лектор

Сенько Олег Валентинович

Лекция 3

Линейная регрессия

Распространённым средством решения задач

прогнозирования величины Y по переменным $X_1, \ldots, X_n$

является использование метода линейной регрессии

$$Y = \beta_0 + \beta_1 X_1 + \ldots \beta_n X_n + \varepsilon$$

Где $\beta_0, \beta_1, \ldots, \beta_n$ регрессионные коэффициенты,

ε - ошибка прогнозирования.

Регрессионные коэффициенты ищутся по обучающей

выборке $\tilde{S}_t = \{(y_1, \mathbf{x}_1), \ldots, (y_m, \mathbf{x}_m)\}$, где $\mathbf{x}_j = (x_{j1}, \ldots, x_{jn})$

- вектор значений переменных $X_1, \ldots, X_n$ для j-го объекта.

Линейная регрессия

Традиционным способом поиска регрессионных коэффициентов

является метод наименьших квадратов (МНК).

МНК заключается в минимизации функционала

$Q(\tilde{S}_t, \beta_0, \ldots, \beta_n) = \frac{1}{m}\sum_{j=1}^{m} [y_j - \beta_0 - \sum_{i=1}^{n} x_{ji}\beta_i]^2$. То есть в качестве

оценок истинных значений регрессионных коэффициентов

берутся значения $\beta_0, \beta_1, \ldots, \beta_n$, для которых $Q(\tilde{S}_t, \beta_0, \ldots, \beta_n)$

Принимает минимальное значение.

Линейная регрессия

Предположим взаимосвязь между величиной Y и переменными $X_1, \ldots, X_n$

описывается выражением ,

$$Y = \beta_0 + \beta_1 X_1 + \ldots \beta_n X_n + \varepsilon_N(0,\sigma) \qquad (1)$$

Где ошибка ε_N распределена нормально, При ‘этом дисперсия ошибки σ^2 не зависит от $X_1, \ldots, X_n$, а математическое ожидание ошибки равно 0 при произвольных значениях прогностических переменных:

$$E_\Omega(\varepsilon_N \mid \mathbf{x}) = 0, \; E_\Omega(\varepsilon_N^2 \mid \mathbf{x}) = \sigma^2 \quad \forall\, \mathbf{x} \in \tilde{X}$$

Линейная регрессия

- В этом случае метод МНК тождественен более общему статистическому методу оценивания параметров статистических распределений – Методу максимального правдоподобия (ММП).
- Метод максимального правдоподобия

Предположим, что некоторое пространство событий, с заданным на нём вероятностной мерой $\mathbf{P}$ характеризуется переменными $Z_1, \ldots, Z_d$

Метод максимального правдоподобия

- Метод ММП позволяет восстанавливать плотность распределения вероятностей по случайным выборкам, если общий вид. плотности вероятностного распределения известен

Пусть плотность распределения $\mathbf{P}$ принадлежит семейству функций, задаваемому вектором параметров $(\theta_1,\ldots,\theta_r)$, принимающем значения из множества $\tilde{\Theta}$

$$\{f(Z_1,\ldots,Z_d,\theta_1,\ldots,\theta_r)\,|\,\boldsymbol{\theta}=(\theta_1,\ldots,\theta_r)\in\tilde{\Theta}\}$$

Метод максимального правдоподобия

Предположим, что у нас имеется случайная выборка объектов, описываемых векторами $\{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$ переменных $Z_1, \ldots, Z_d$

Метод МП заключается в выборе в семействе $\{f(Z_1, \ldots, Z_d, \theta_1, \ldots, \theta_r) \mid \boldsymbol{\theta} \in \tilde{\Theta}\}$ плотности, для которой достигает максимума функция правдоподобия

$$L(\theta_1, \ldots, \theta_r) = \prod_{j=1}^{m} f(\mathbf{z}_j, \boldsymbol{\theta})$$

Метод максимального правдоподобия

Иными словами оценка $\hat{\boldsymbol{\theta}}$ вектора параметров

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$ вычисляется как

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \tilde{\Theta}} \{ L(\mathbf{z}_1, \ldots, \mathbf{z}_m, \theta_1, \ldots, \theta_r) \}$$

- Согласно модели (1) разность $Y - \beta_0 - \beta_1 X_1 - \ldots - \beta_n X_n$

Подчиняется нормальному распределению с нулевым

математическим ожиданием и дисперсией σ^2

Соответствие ММП и МНК

Плотность распределения в пространстве переменных

$(Y, X_1, \dots, X_n)$ может быть восстановлена по обучающей

выборке $\tilde{S}_t = \{(y_1, \mathbf{x}_1), \dots, (y_m, \mathbf{x}_m)\}$

путём максимизации функции правдоподобия

$$L(\tilde{S}_t, \beta_0, \dots, \beta_0) = \prod_{j=1}^{m} \frac{1}{(\sqrt{2\pi\sigma})^m} \exp[\frac{-(y_j - \beta_0 - \sum_{i=1}^{n} x_{ji}\beta_i)^2}{2\sigma^2}]$$

Соответствие ММП и МНК

Очевидно, точка экстремума функции правдоподобия

$L(\tilde{S}_t, \beta_0, \ldots, \beta_n)$ совпадает с точкой экстремума функции

$\ln[L(\tilde{S}_t, \beta_0, \ldots, \beta_n)]$

$$\ln[L(\tilde{S}_t, \beta_0, \ldots, \beta_n)] = \sum_{j=1}^{m} [-\tfrac{1}{2}\ln(2\pi) - \ln(\sigma)] +$$

$$-\tfrac{1}{2\sigma^2}\sum_{j=1}^{m}(y_j - \beta_0 - \sum_{i=1}^{n} x_{ji}\beta_i)^2$$

Очевидно, что точка максимума $\ln[L(\tilde{S}_t, \beta_0, \ldots, \beta_n)]$ совпадает с точкой минимума функции $Q(\tilde{S}_t, \beta_0, \ldots, \beta_n)$, оптимизируемой в методе МНК, , что позволяет сделать вывод о эквивалентности ММП и МНК

Одномерная линейная регрессия

Метод одномерной регрессии позволяет восстановит линейную зависимость переменной Y от единственной переменной X по обучающей выборке $\tilde{S}_t = \{(y_1, x_1), \ldots, (y_m, x_m)\}$

МНК заключается в минимизации функционала

$$Q(\tilde{S}_t, \beta_0, \beta_1) = \tfrac{1}{m} \sum_{j=1}^{m} [y_j - \beta_0 - \beta_1 x_j]^2$$

Иными словами оценки истинных значений β - параметров

$(\hat{\beta}_0, \hat{\beta}_1)$ вычисляются как

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0, \beta_1) \in \tilde{\mathrm{B}}}{\arg\min} \tfrac{1}{m} \sum_{j=1}^{m} [y_j - \beta_0 - \beta_1 x_j]^2$$

Одномерная линейная регрессия

Необходимым условием минимума функционала $Q(\tilde{S}_t, \beta_0, \beta_1)$

является выполнение системы из двух уравнений

$$\frac{\partial Q(\tilde{S}_t, \beta_0, \beta_1)}{\partial \beta_0} = \frac{2}{m}\sum_{j=1}^{m} y_j - 2\beta_0 - \frac{2\beta_1}{m}\sum_{j=1}^{m} x_j = 0$$

$$\frac{\partial Q(\tilde{S}_t, \beta_0, \beta_1)}{\partial \beta_1} = \frac{2}{m}\sum_{j=1}^{m} x_j y_j - \frac{2\beta_0}{m}\sum_{j=1}^{m} x_j - \frac{2\beta_1}{m}\sum_{j=1}^{m} x_j^2 = 0 \qquad (2)$$

Оценки $(\hat{\beta}_0, \hat{\beta}_1)$ являются решением системы неравенств (2)

относительно параметров (β_0, β_1) соответственно

Одномерная линейная регрессия

Таким образом оценки могут быть записаны в виде

$$\hat{\beta}_1 = \frac{\sum_{j=1}^{m} x_j y_j - \frac{1}{m}\sum_{j=1}^{m} x_j \sum_{j=1}^{m} y_j}{\sum_{j=1}^{m} x_j^2 - \frac{1}{m}(\sum_{j=1}^{m} x_j)^2},$$

$$\hat{\beta}_0 = \overline{y} - \beta_1 \overline{x}$$, где

$$\overline{y} = \frac{1}{m}\sum_{j=1}^{m} y_j, \quad \overline{x} = \frac{1}{m}\sum_{j=1}^{m} x_j$$

Выражение для $\hat{\beta}_1$ может быть переписано в виде

$$\hat{\beta}_1 = \frac{Cov(Y, X \mid \tilde{S}_t)}{D(X)},$$ где $$Cov(Y, X \mid \tilde{S}_t) = \frac{1}{m}\sum_{j=1}^{m}(y_j - \overline{y})(x_j - \overline{x})$$

$$D(X \mid \tilde{S}_t) = \frac{1}{m}\sum_{j=1}^{m}(x_j - \overline{x})^2$$ соответственно

выборочные ковариация и дисперсия

Многомерная линейная регрессия

При вычислении оценки вектора β - параметров в случае многомерной линейной регрессии удобно использовать матрицу плана $\mathbf{X}$ размера $m \times (n+1)$, которая строится по обучающей выборке

$$\tilde{S}_t = \{(y_1, \mathbf{x}_1), \ldots, (y_m, \mathbf{x}_m)\}$$

$$\mathbf{X} = \begin{pmatrix} 1\, x_{11} \ldots\ldots\ldots x_{1n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ 1\, x_{j1} \ldots\ldots\ldots x_{jn} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ 1\, x_{m1} \ldots\ldots\ldots x_{mn} \end{pmatrix}$$

Многомерная линейная регрессия

Пусть $\mathbf{y} = (y_1, \ldots, y_m)$ - вектор значений переменной Y.

Связь значений Y с переменными $(X_1, \ldots, X_n)$ на объектах обучающей выборки может быть описана с помощью матричного уравнения $\mathbf{y} = \boldsymbol{\beta} \mathbf{X}^t + \boldsymbol{\varepsilon}$, где

$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_m)$ - вектор ошибок для объектов $\tilde{S}_t$.

Функционал $Q(\tilde{S}_t, \beta_0, \ldots, \beta_n)$ Может быть записан в виде

$Q(\tilde{S}_t, \beta_0, \ldots, \beta_n) = \frac{1}{m} \sum_{j=1}^{m} [y_j - \sum_{i=1}^{n+1} \beta_i \widehat{x}_{ji}]^2$, где $\widehat{x}_{ji}$ - элемент $\mathbf{X}$

Многомерная линейная регрессия

Необходимым условием минимума функционала $Q(\tilde{S}_t, \beta_0, \ldots, \beta_n)$ является выполнение системы из $n+1$ уравнений

$$\frac{\partial Q(\tilde{S}_t, \beta_0, \ldots, \beta_n)}{\partial \beta_0} = 2[\sum_{j=1}^{m} y_j \widehat{x}_{j1} - \sum_{j=1}^{m} \sum_{i=1}^{n+1} \beta_i \widehat{x}_{ji} \widehat{x}_{j1}] = 0 \quad (3)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\frac{\partial Q(\tilde{S}_t, \beta_0, \ldots, \beta_n)}{\partial \beta_n} = 2[\sum_{j=1}^{m} y_j \widehat{x}_{jn} - \sum_{j=1}^{m} \sum_{i=1}^{n+1} \beta_i \widehat{x}_{ji} \widehat{x}_{jn}] = 0$$

Многомерная линейная регрессия

В матричной форме система (3) может быть записана в виде

$$-2\mathbf{X}^t\mathbf{y}^t + 2\mathbf{X}^t\mathbf{X}\boldsymbol{\beta}^t = 0 \qquad (4)$$

Вектор Оценок истинных значений регрессионных коэффициентов $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_n)$ является решением системы уравнений (4) относительно параметров соответственно. $(\beta_0, \ldots, \beta_n)$

Решение системы (4) существует, если $\det(\mathbf{X}^t\mathbf{X}) \neq 0$

Многомерная линейная регрессия

- В этом случае для $\mathbf{X}^t\mathbf{X}$ существует обратная матрица и решение (4) относительно вектора может быть записано в виде: $\hat{\boldsymbol{\beta}}^t = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t y^t$

МУЛЬТИКОЛЛИНЕАРНОСТЬ

Явление мультиколлинеарности,

Из теории матриц следует, что $\det(\mathbf{X}^t\mathbf{X}) = 0$ если ранг матрицы $\mathbf{X}$ по строкам менее n. Однако при сильной коррелированности одной из переменной с какой-либо линейной комбинацией других переменных

значение $\det(\mathbf{X}^t\mathbf{X})$ оказывается близким 0

При этом вычисленный вектор оценок $\hat{\boldsymbol{\beta}}^t$ может сильно изменяться при небольших изменениях в обучающей выборке..

Свойства оптимальных регрессий

- Рассмотрим свойства линейных регрессий, минимизирующих квадрат ошибки на пространстве событий Ω. Пусть $R(X_1,\ldots,X_n)$ - регрессионная функция, которая не может быть улучшена с помощью дополнительного линейного преобразования. Иными словами
- $$\forall \alpha_0,\alpha_1 \quad E_\Omega(Y-\alpha_0-\alpha_1 R)^2 \ge E_\Omega(Y-R)^2$$

Свойства оптимальных регрессий

- То есть минимум $E_\Omega(Y-\alpha_0-\alpha_1 R)^2$ достигается при

$$\alpha_0=0, \quad \alpha_1=1$$

$$E_\Omega(Y-\alpha_0-\alpha_1 R)^2 = E_\Omega Y^2 - 2\alpha_0 E_\Omega Y -$$

$$-2\alpha_1 E_\Omega(YR) + \alpha_1^2 E_\Omega R^2 + 2\alpha_1\alpha_0 E_\Omega R + \alpha_0^2$$

Необходимым условием экстремума $E_\Omega(Y-\alpha_0-\alpha_1 R)^2$

является равенство 0 частных производных

$$\frac{\partial E_\Omega(Y-\alpha_0-\alpha_1 R)^2}{\partial \alpha_0}, \quad \frac{\partial E_\Omega(Y-\alpha_0-\alpha_1 R)^2}{\partial \alpha_1}$$

Свойства оптимальных регрессий

Что эквивалентно уравнениям

$$2\alpha_1 E_\Omega R + 2\alpha_0 - 2\alpha_1 E_\Omega Y = 0$$

$$-2E_\Omega(YR) + 2\alpha_1 E_\Omega R^2 + 2\alpha_0 E_\Omega R = 0$$

Принимая во внимание, что в точке экстремума $\alpha_0 = 0, \quad \alpha_1 = 1$

получаем следующие свойства оптимального линейного

прогнозирующего алгоритма

1) $E_\Omega R = E_\Omega Y$ 2) $E_\Omega R^2 = E_\Omega (YR)$

Свойства оптимальных регрессий

- Из свойств 1) 2) следует , что дисперсия R

равна ковариации Y и R

$$D(R) = E_\Omega (R - E_\Omega R)^2 = E_\Omega R^2 - (E_\Omega R)^2$$

$$\operatorname{cov}(YR) = E_\Omega \{(R - E_\Omega R)(Y - E_\Omega Y)\} = E_\Omega (RY) - (E_\Omega R)^2$$

То есть $3) \operatorname{cov}(YR) = D(R)$

Свойства оптимальных регрессий

Рассмотрим коэффициент корреляции между Y и R

$$3)\quad K(YR)=\frac{\operatorname{cov}(YR)}{\sqrt{D(Y)D(R)}}=\sqrt{\frac{D(R)}{D(Y)}}$$

Величина ошибки прогнозирования Y с помощью R

$$4)\Delta(Y,R)=E_{\Omega}(Y-R)^2=E_{\Omega}Y^2-2E_{\Omega}(YR)+E_{\Omega}R^2=$$

$$=E_{\Omega}Y^2-E_{\Omega}R^2=E_{\Omega}Y^2-(E_{\Omega}Y)^2+(E_{\Omega}Y)^2-E_{\Omega}R^2=$$

$$=E_{\Omega}Y^2-(E_{\Omega}Y)^2+(E_{\Omega}R)^2-E_{\Omega}R^2=D(Y)-D(R)$$

Свойства оптимальных регрессий

Из свойств (3) и (4) легко следует свойство для

относительной ошибки $\Delta_r(Y,R) = \Delta_r(Y,R)/D(Y)$

$$5)\Delta_r(Y,R) = 1 - K^2(Y,R)$$

Разложение обобщённой ошибки

Напомним, что обобщающая способность алгоритма прогнозирования $A(\mathbf{x},\tilde{S}_t)$, обученного по выборке $\tilde{S}_t$ с помощью метода $\mathbf{A}$ измеряется величиной потерь на генеральной совокупности Ω

$$E_\Omega\{\lambda[Y,A(\mathbf{x},\tilde{S}_t)]\}=\int_\Omega \lambda[Y,A(\mathbf{x})]P(d\omega)$$

Разложение обобщённой ошибки

- Для оценки эффективности использования метода прогнозирования $\mathbf{A}$ для прогнозирования случайного процесса, связанного с генеральной совокупностью Ω при размере обучающей выборки m естественно использовать математическое ожидание потерь по пространству всевозможных обучающих выборок $\tilde{S}_m$ длины - m $\Omega_m = \Omega \times \ldots \times \Omega$

$$E_{\Omega_m} E_{\Omega}\{\lambda[Y, A(\mathbf{x}, \tilde{S}_m)]\}$$

Разложение обобщённой ошибки

При использовании в качестве функции потерь квадрата ошибки $\lambda[y_j, A(\mathbf{x}_j)] = [y_j - A(\mathbf{x}_j)]^2$ обобщённые потери (обобщённая квадратичная ошибка Δ_G) принимает вид

$$\Delta_G = E_{\Omega_m} E_{\Omega}\{[Y - A(\mathbf{x}, \tilde{S}_m)]^2\}$$

Проведём преобразования

$$\Delta_G = E_{\Omega_m} E_{\Omega}\{[Y - E(Y \mid \mathbf{x}) + E(Y \mid \mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\} =$$

$$= E_{\Omega_m} E_{\Omega}\{[Y - E(Y \mid \mathbf{x})]\}^2 + E_{\Omega_m} E_{\Omega}\{[E(Y \mid \mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\} +$$

$$+ E_{\Omega_m} E_{\Omega}\{[E(Y \mid \mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)][Y - E(Y \mid \mathbf{x})]\}$$

Разложение обобщённой ошибки

Справедливо равенство

$$E_{\Omega_m} E_{\Omega}\{[E(Y\,|\,\mathbf{x}) - A(\mathbf{x},\tilde{S}_m)][Y - E(Y\,|\,\mathbf{x})]\} = 0$$,

которое следует из того,

что для при любом $\mathbf{x}$ $E_{\Omega_{\mathbf{x}}}\{[Y - E(Y\,|\,\mathbf{x})]\} = 0$

Принимая во внимание, что $[Y - E(Y\,|\,\mathbf{x})]^2$ не зависит от $\tilde{S}_m$

получаем $E_{\Omega_m} E_{\Omega}\{[Y - E(Y\,|\,\mathbf{x})]^2\} = E_{\Omega}\{[Y - E(Y\,|\,\mathbf{x})]^2\}$

Разложение обобщённой ошибки

В итоге

$$\Delta_G = E_\Omega\{[Y - E(Y\,|\,\mathbf{x})]^2\} + E_{\Omega_m} E_\Omega[E(Y\,|\,\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\}$$

Введём обозначение

$$\hat{A}(\mathbf{x}) = E_{\Omega_m}\{A(\mathbf{x}, \tilde{S}_m)\}$$

Компонента разложения

$$E_{\Omega_m} E_\Omega[E(Y\,|\,\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\}$$

Может быть представлена в виде

$$E_{\Omega_m} E_\Omega[E(Y\,|\,\mathbf{x}) - \hat{A}(\mathbf{x}) + \hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\} =$$

Разложение обобщённой ошибки

$$= E_{\Omega_m} E_{\Omega}\{[E(Y\,|\,\mathbf{x}) - \hat{A}(\mathbf{x})]^2\} + E_{\Omega_m} E_{\Omega}\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\} +$$

$$+ E_{\Omega_m} E_{\Omega}\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)][E(Y\,|\,\mathbf{x}) - \hat{A}(\mathbf{x})]\}$$

Справедливо равенство

$$E_{\Omega_m} E_{\Omega}\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)][E(Y\,|\,\mathbf{x}) - \hat{A}(\mathbf{x})]\} = 0$$

Действительно

$$E_{\Omega_m} E_{\Omega}\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)][E(Y\,|\,\mathbf{x}) - \hat{A}(\mathbf{x})]\} =$$

$$= E_{\Omega}\{[E(Y\,|\,\mathbf{x}) - \hat{A}(\mathbf{x})] E_{\Omega_m}\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]\}\}$$

Разложение обобщённой ошибки

Из определения $\hat{A}(\mathbf{x})$ следует

$$E_{\Omega_m}\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]\} = 0$$

В итоге справедливо трёхкомпонентное разложение обобщённой квадратичной ошибки Δ_G

$$\Delta_G = E_\Omega\{[Y - E(Y \mid \mathbf{x})]^2\} + E_\Omega\{[E(Y \mid \mathbf{x}) - \hat{A}(\mathbf{x})]^2\} +$$

$$+ E_{\Omega_m} E_\Omega\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\} =$$

$$= \Delta_N + \Delta_B + \Delta_V$$

Разложение обобщённой ошибки

Шумовая компонента

$$\Delta_N = E_\Omega\{[Y - E(Y\,|\,\mathbf{x})]^2\}$$

является минимально достижимой квадратичной ошибкой прогноза, которая не может быть устранена с использованием только математических средств.

Разложение обобщённой ошибки

Составляющая сдвига (Bias)

$$\Delta_B = E_\Omega\{[E(Y \mid \mathbf{x}) - \hat{A}(\mathbf{x})]^2\}$$

Высокое значение компоненты сдвига в модели $\tilde{M} = \{A : \tilde{X} \to \tilde{Y}\}$

Алгоритмов, достаточно хорошо аппроксимирующих объективно существующую зависимость Y от переменных $X_1, \ldots, X_n$

Составляющая сдвига может быть снижена путём включения в модель

Дополнительных алгоритмов прогнозирования, позволяющих повысить точность аппроксимации

Разложение обобщённой ошибки

Дисперсионная составляющая (Variance)

$$\Delta_V = E_{\Omega_m} E_{\Omega}\{[\hat{A}(\mathbf{x}) - A(\mathbf{x}, \tilde{S}_m)]^2\}$$

характеризует неустойчивость обученных прогнозирующих алгоритмов при статистически возможных изменениях в обучающих выборках. Дисперсионная составляющая возрастает при небольших размерах обучающей выборки. Дисперсионная составляющая может быть снижена путём выбора сложности модели, соответствующей размеру обучающих данных.

Разложение обобщённой ошибки

Таким образом существует

Bias-Variance дилемма

Составляющая сдвига может быть снижена путём увеличения разнообразия модели. Однако увеличение разнообразия модели при недостаточном объёме обучающих данных ведёт к росту компоненты сдвига.

Наиболее высокая точность прогноза достигается, при поддержании правильного баланса между разнообразием используемой модели и объёмом обучающих данных