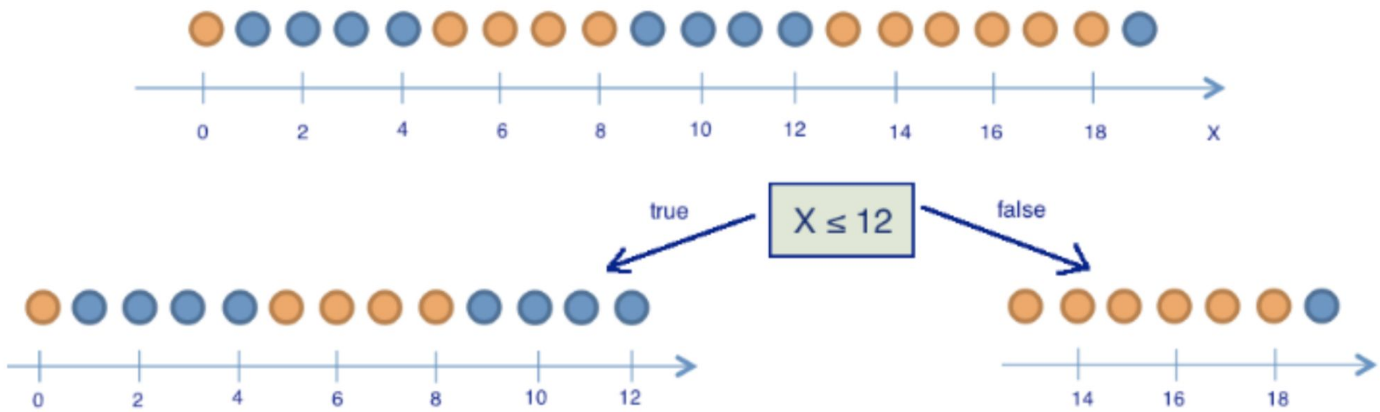
Machine Learning

myTracker

mail.ru group

Линейные модели

Деревья решений (регрессия)



Критерий. Выборочное среднее

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2$$

Ответ. Среднее значений в листе

$$X_{leaf} = \frac{1}{n}\sum_{i=1}^{n}X_i$$

Метод наименьших квадратов

Определение. Пусть задана такая зависимость: $y_t = f(x_t, b) + \varepsilon_t$, где ϵ_t - случайная ошибка модели и b - набор неизвестных параметров. Надо восстановить изначальную зависимость y от x. Для этого подберем параметры b наилучшим образом.

Определение. Введем функцию “ошибки”, с помощью которой будем оценивать параметры b

$$RSS(b) = e^T e = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (y_t - f(x_t, b))^2$$

Задача. Найти $\hat{b}_{OLS} = \arg\min_b RSS(b)$

Метод наименьших квадратов (линейная модель)

Определение. Пусть задана линейная зависимость

$$y_t = \sum_{j=1}^{k} b_j x_{tj} + \varepsilon = x_t^T b + \varepsilon_t \quad <-> \quad y = Xb + \varepsilon.$$

Функциональное представление

Матричное представление

Определение. Функция ошибки в матричном представлении имеет вид

$$RSS = e^T e = (y - Xb)^T (y - Xb)$$

Если продифференцировать по вектору параметров b и приравняем производную к нулю, получаем

$$(X^T X)b = X^T y.$$

Регуляризация

В итоге решение линейной регрессии можно найти по формуле:

$$b = (X^T X)^{-1} X^T y$$

Регуляризация

В итоге решение линейной регрессии можно найти по формуле:

$$b = (X^T X)^{-1} X^T y$$

Проблема. Матрица $X^T X$ становится необратимой.

Решение. Задачу надо изменить и сделать матрицу обратимой или регулярной.

Регуляризация

В итоге решение линейной регрессии можно найти по формуле:

$$b = (X^T X)^{-1} X^T y$$

Проблема. Матрица $X^T X$ становится необратимой.

Решение. Задачу надо изменить и сделать матрицу обратимой или регулярной.

Проблема. Матрица с мультиколлинарными столбцами дает нестабильную оценку параметров.

Решение. Добавить ограничение на параметры.

$$Error = RSS + \lambda b^T b$$

Регуляризация

В итоге решение линейной регрессии можно найти по формуле:

$$b = (X^T X)^{-1} X^T y$$

Проблема. Матрица $X^T X$ становится необратимой.

Решение. Задачу надо изменить и сделать матрицу обратимой или регулярной.

Проблема. Матрица с мультиколлинарными столбцами дает нестабильную оценку параметров.

Решение. Добавить ограничение на параметры.

$$Error = RSS + \lambda b^T b$$

В итоге получаем гребневую регрессию (Ridge regression):

$$b = (X^T X + \lambda E)^{-1} X^T y$$

Линейные регрессии с регуляризациями

Ridge regression (гребневая регрессия, регрессия с $L2$ регуляризацией)

$$Error = RSS + \lambda \sum_i b_i^2$$

LASSO regression (лассо регрессия, регрессия с $L1$ регуляризацией)

$$Error = RSS + \lambda \sum_i |b_i|$$

Можно комбинировать регуляризации (Elastic Net regression)

$$Error = RSS + \lambda_1 \sum_i |b_i| + \lambda_2 \sum_i b_i^2$$

Линейные регрессии с регуляризациями

Ridge regression (гребневая регрессия, регрессия с $L2$ регуляризацией)

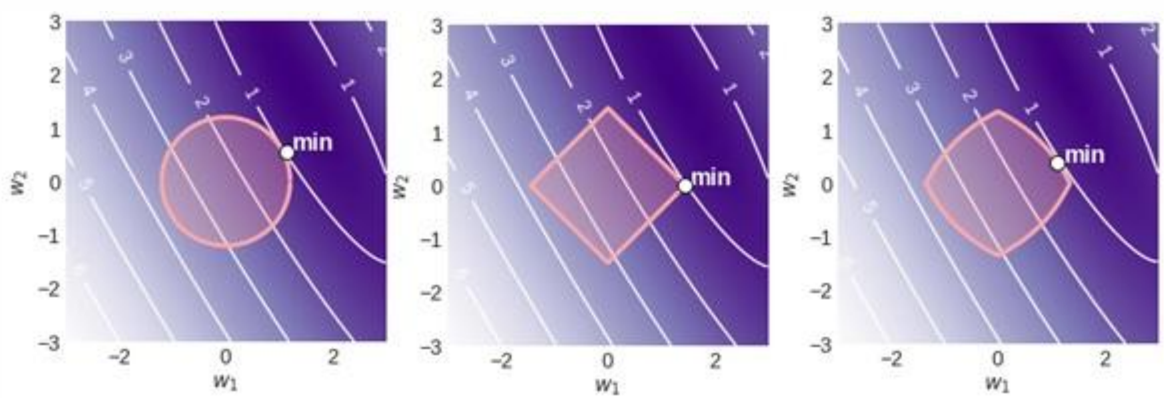
$$Error = RSS + \lambda \sum_i b_i^2$$

LASSO regression (лассо регрессия, регрессия с $L1$ регуляризацией)

$$Error = RSS + \lambda \sum_i |b_i|$$

Можно комбинировать регуляризации (Elastic Net regression)

$$Error = RSS + \lambda_1 \sum_i |b_i| + \lambda_2 \sum_i b_i^2$$



L_2 $\quad$ L_1 $\quad$ $L_1 + L_2$

Логистическая регрессия

Определение. Логит-функцией (сигмоидой) назовем функцию вида:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Определение. Вероятностью отнесения объекта к положительному классу будем рассчитывать по формуле:

$$p_+(x) = P(y = 1|x, b) = \sigma(b^T x)$$

Логистическая регрессия

Определение. Логит-функцией (сигмоидой) назовем функцию вида:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Определение. Вероятностью отнесения объекта к положительному классу будем рассчитывать по формуле:

$$p_+(x) = P(y=1|x,b) = \sigma(b^T x)$$

Определение. Вероятностью отнесения объекта к отрицательному классу будем рассчитывать по формуле:

$$p_-(x) = P(y=-1|x,b) = 1 - \sigma(b^T x) = \sigma(-b^T x)$$

Определение. Вероятностью отнесения объекта к своему классу будем рассчитывать по формуле:

$$p(x) = \sigma(yb^T x)$$

Логистическая регрессия

Вопрос. Как обучать такую модель?

Ответ. Воспользуемся методом максимального правдоподобия.

Логистическая регрессия

Вопрос. Как обучать такую модель?

Ответ. Воспользуемся методом максимального правдоподобия.

Решение.

1. Построим функцию максимального правдоподобия

$$L = \prod p(x_i | \theta) = \prod_i \sigma(y_i b^T x_i)$$

Логистическая регрессия

Вопрос. Как обучать такую модель?

Ответ. Воспользуемся методом максимального правдоподобия.

Решение.

1. Построим функцию максимального правдоподобия

$$L = \prod p(x_i|\theta) = \prod_i \sigma(y_i b^T x_i)$$

2. Применим логарифм к функции правдоподобия

$$\log L = \log \prod_i \sigma(y_i b^T x_i) = \sum_i \log \sigma(y_i b^T x_i) = -\sum_i \log(1 + e^{y_i b^T x_i})$$

Логистическая регрессия

Вопрос. Как обучать такую модель?

Ответ. Воспользуемся методом максимального правдоподобия.

Решение.

1. Построим функцию максимального правдоподобия

$$L = \prod p(x_i|\theta) = \prod_i \sigma(y_i b^T x_i)$$

2. Применим логарифм к функции правдоподобия

$$\log L = \log \prod_i \sigma(y_i b^T x_i) = \sum_i \log \sigma(y_i b^T x_i) = -\sum_i \log(1 + e^{y_i b^T x_i})$$

3. Получаем следующий функционал, который надо минимизировать:

$$Error = \sum_i \log(1 + e^{y_i b^T x_i})$$