Лекция 2. Линейная регрессия.

Рассматриваем: $\mathbb{Y} = \mathbb{R}$, $\mathbb{X} = \mathbb{R}^2 \Rightarrow$ все признаки вещественные.

Линейная модель: $a(x) = \underbrace{w_0}_{bias} + \sum_{j=1}^{d} \underbrace{w_j}_{weights} \cdot x_j$

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d$, $w = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^d$, тогда

$a(x) = w_0 + < w, x >$

Предроложение: есть константный признак $x_1 \equiv 1 \quad \Rightarrow \quad \ldots + w_1 x_1 + \cdots = \cdots w_1 + \cdots$, тогда

$a(x) = < w, x >$

Применимость линейных моделей:

$a(x) = w_0 + w_1 \cdot \text{площадь} + w_2 \cdot \text{этаж} + w_3 \cdot \text{расстояние до метро} \Rightarrow$ ⋆ признаки влияют независимо ⋆ признаки влияют линейно $\Rightarrow$ для линейных моделей данные нужно готовить!

Функции потерь:

MSE. $L(y,z) = (y-z)^2$, $Q(a,X) = \frac{1}{\ell}\sum_{i=1}^{\ell}(a(x_i) - y_i)^2$,

RMSE. $Q(a,X) = \sqrt{\frac{1}{\ell}\sum_{i=1}^{\ell}(a(x) - y_i)^2}$

(1. похоже на дисперсию. 2. стимулирует подгонку под выбросы. 3. по производной можно понять близость к экстремуму.)

$\underbrace{R^2}_{\text{коэффициент детерминации}} = 1 - \frac{\sum_{i=1}^{\ell}(a(x_i) - y_i)^2}{\sum_{i=1}^{\ell}(y_i - \overline{y})^2}$, $a(x_i) = y_i \Rightarrow R^2 = 1$, $a(x_i) = \overline{y} \Rightarrow R^2 = 0$

То есть для разумных моделей $\boxed{0 < R^2 \leq 1}$, чем ближе к единице, тем лучше!

MAE. $L(y,z) = |y - z|$, $Q(a,X) = \frac{1}{\ell}\sum_{i=1}^{\ell}|a(x_i) - y_i|$

(1. лучше игнорирует выбросы. 2. производная везде одинаковая (± 1), что не полезно.)

Huber loss. $L_\delta(y,z) = \begin{cases} \frac{1}{2}(y-z)^2, & |y-z| < \delta \\ \delta\left(|y-z| - \frac{1}{2}\delta\right), & |y-z| \geq \delta \end{cases}$

(1. при квадратичном поведении: сильно штрафуем за отклонения от правильного отклонения. 2. при линейном поведении: все равно на большие ошибки. 3. Итог: чем больше δ, тем более сильные ошибки не считаем за выбросы. 4. проблема со второй производной.)

Log-Cosh. $L(y,z) = \log\cosh(y - z)$

(1. вторая производная непрерывна. 2. точка разрыва в точке перехода.)

MSLE. $L(y,z) = (\log(z+1) - \log(y+1))^2$, $\begin{cases} y \geq 0 \\ z \geq 0 \end{cases}$

Относительные функции потерь. $L(y,z) = \left|\frac{y-z}{y}\right|$, $Q(a,X) = \frac{1}{\ell}\sum_{i=1}^{\ell}\left|\frac{y_i - a(x_i)}{y_i}\right| - MAPE$

(1. Интерпретируемость. $L = 2 \Rightarrow$ ошибся в 3 раза. 2. Хороши при разных масштабах y. $y_1, y_2 = 1000{,}1$ $z_1, z_2 = 999{,}0$)

Квантильная функция потерь. $L(y,z) = (\tau - 1)[y - z < 0](y - z) + \tau[y - z \geq 0](y - z)$