L2

Recap:

① representation

② search:

Document ⇒ vector of term frequency

$x_{i(j)} = tf_{i(j)}$ "Term Frequency"

$x_{i(j)} = tf_{i(j)} \cdot \underbrace{idf_{(j)}}$ → "Inverse Document Frequency"

$idf_{(j)} = \log\left(\frac{n}{n_j}\right)$

n → number of documents

n_j → number of documents containing word n_j

Pros	Cons
- simple	- $O(nd)$ time per test instance
- robust	

Regression

Dataset (X_i, Y_i) — data (X_i), label (Y_i)

⋮ ⋮

(X_n, Y_n)

goal of regression: Find a function of $y_i \approx f(X_i)$

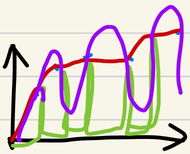
Application: ① Classification of Image:

x_i image $y_i \to$ cat / no cat (类别)

② Retail Pricing: x: time y: price of product

③ weather:

X: location y: rainfall

假设f bad →  many function to map x to y

Not well posed → Not robust since there are a lot of f

Fix Regression: find a function f that belongs to a class of functions H 固定一类映射函数

寻找好的线性模型

Hypothesis class

Linear models:

$\mathcal{H}$: set of linear functions

change in input $\propto$ change in output

why linear models?

- simplicity
- stable behaviour
- easy to compute
- Interpretable
  高维、清晰
- building block of more complex function

$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots$

泰勒展开，线性模拟.

linear regression (univariate)

x → scalar
y → scalar

(x_1, y_1)
(x_2, y_2)
$\vdots \quad \vdots$
(x_n, y_n)

$\hat{y}$ 表示由x对y的估计

Step 1: Representation

$$y = w_0 + w_1 x$$

$$\hat{y}_i = w_0 + w_1 x_i$$

Step 2: measure of goodness :

eg. MSE (mean square error)

$$E_y(\hat{y} - y)^2$$

测量值

predicted value

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(w_0 + w_1 x_i - y)^2$$

算出 w_0, w_1 得出线性映射关系

Step 3: find best w_0, w_1 that minimize MSE

take partial derivative over w_0, w_1

偏导

① $\frac{\partial MSE}{\partial w_0} = 0$

寻找最小MSE对应的系数 w_0, w_1

$$2 \cdot \frac{1}{n}\sum_{i=1}^{n}(w_0 + w_1 x_i - y_i) = 0$$

solve for w_0

$$w_0 = \frac{\sum_i y_i}{n} - w_1 \frac{\sum x_i}{n}$$

$$w_0 = \bar{y} - w_1 \bar{x}$$

② $\frac{\partial MSE}{\partial w_1} = 0$

$$2 \cdot \frac{1}{n}\sum_{i=1}^{n} x_i (w_0 + w_1 x_i - y_i) = 0$$

$$\frac{1}{n} \cdot \sum x_i (\bar{y} - w_1 \bar{x} + w_1 x_i - y_i) = 0$$

$$w_1 \cdot \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) x_i = \frac{1}{n}\sum (y_i - \bar{y}) x_i$$

协方差

$$w_1 = \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum (x_i - \bar{x})^2}$$

方差

$Cov(X, Y)$
$= E[(X - \bar{X})(Y - \bar{Y})]$
$= E(XY) - E(X)E(Y)$

$$w_1 = \frac{\left(\frac{1}{n}\sum x_i y_i\right) - (\bar{x}\bar{y})}{\frac{1}{n}\sum x_i^2 - \bar{x}^2}$$

协方差

cross-covariance of X & y

variance of X

Variance

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$=\frac{1}{n}\sum\left(x_i^2+\bar{x}^2-2x_i\bar{x}\right)$$

$$=\boxed{\frac{1}{n}\sum x_i^2-\bar{x}^2}$$

$\sigma_x^2 \rightarrow$ variance of x

$\sigma_y^2 \rightarrow$ variance of y

$\sigma_{xy} \rightarrow$ cross-covariance of x & y

测量值

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left[\underline{y_i}-\underline{(\overbrace{\bar{y}-w_1\bar{x}}^{w_0}+w_1x_i)}\right]^2 \rightarrow \hat{y}$$

$$=\frac{1}{n}\sum\left[(y_i-\bar{y})-\underbrace{\frac{\sigma_{xy}}{\sigma_x^2}}_{w_1}(x_i-\bar{x})\right]^2$$

$$=\frac{1}{n}\sum(y_i-\bar{y})^2+\frac{1}{n}\sum\left(\frac{\sigma_{xy}}{\sigma_x^2}\right)^2(x_i-\bar{x})^2-2\cdot\frac{1}{n}\sum\frac{\sigma_{xy}(y_i-\bar{y})(x_i-\bar{x})}{\sigma_x^2}$$

$$=\sigma_y^2+\frac{\sigma_{xy}^2}{\sigma_x^4}\cdot\sigma_x^2-2\cdot\frac{\sigma_{xy}^2}{\sigma_x^2}$$

FVU.

$$=\sigma_y^2+\frac{\sigma_{xy}^2}{\sigma_x^2}-\frac{2\sigma_{xy}^2}{\sigma_x^2}$$

$$\frac{E(\hat{y}-y)^2}{Var(y)}$$

$$MSE=\sigma_y^2-\frac{\sigma_{xy}^2}{\sigma_x^2}$$

Fraction of Variance Unexplained

(MSE=0)

$$\text{ideal } 0=\frac{MSE}{\sigma_y^2}=1-\frac{\sigma_{xy}^2}{\sigma_x^2\sigma_y^2}$$

R^2 value

ideal : $R^2=1$

measure how good model is

$$\frac{MSE}{\sigma_y^2}=0 \quad (MSE=0)$$