Linear Regression:

Jun 2, 2021

$$\hat{y} = mx + b$$

\# features

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{n}\left(y_i - (wx_i + b)\right)^2 \quad \text{MSE}$$

\# examples

$$J'(w,b) = \begin{bmatrix} \frac{df}{dw} \\ \frac{db}{dw} \end{bmatrix} = \begin{bmatrix} \frac{1}{m}\sum_{i=1}^{n} -2x_i\left(y_i - \overbrace{(wx_i + b)}^{\hat{y}}\right) \\ \frac{1}{m}\sum_{i=1}^{n} -2\left(y_i - (wx_i + b)\right) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{m}\sum_{i=1}^{n} \cancel{2}x_i(\hat{y} - y_i) \\ \frac{1}{m}\sum_{i=1}^{n} \cancel{2}(\hat{y} - y_i) \end{bmatrix}$$

↳ remove constants cuz it's supposed to be $\frac{1}{2m}$ anyways

$$w = w - \alpha \cdot dw$$
$$b = b - \alpha \cdot db$$

Logistic Regression:

$$z = wx + b$$

$$\hat{y} = \text{sigmoid}(z) = \frac{1}{1+e^{-z}}$$

$$\therefore \hat{y} = h_\theta(x) = \frac{1}{1+e^{-wx+b}}$$

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{n}\left[y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right]$$

$$J'(w,b) = \begin{bmatrix} \frac{1}{m}\sum_{i=1}^{n} x_i(\hat{y}-y_i) \\ \frac{1}{m}\sum_{i=1}^{n} (\hat{y}-y_i) \end{bmatrix}$$

$$w = w - \alpha \cdot dw$$

$$b = b - \alpha \cdot db$$

Support Vector Machines:

$$\left.\begin{aligned} w\cdot x_i - b \geq 1 & \text{ if } y_i = 1 \\ w\cdot x_i - b \leq -1 & \text{ if } y_i = -1 \end{aligned}\right\} \therefore y_i(w\cdot x_i - b) \geq 1$$

Loss:

$$l = \begin{cases} 0 & \text{if } y_i(w\cdot x_i - b) \geq 1 \\ 1 - y_i(w\cdot x_i - b) & \text{otherwise} \end{cases}$$

$\hookrightarrow l = \max(0, 1 - y_i(w\cdot x_i - b))$

Regularization:

$$J = \frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - y_i(w\cdot x_i - b)) + \lambda \|w\|^2$$

$\hookrightarrow$ if $y_i(w\cdot x_i - b) \geq 1$:

$$J_i = \lambda \|w\|^2$$

$$\frac{dJ_i}{dw} = 2\lambda w, \quad \frac{dJ_i}{db} = 0$$

else:

$$J_i = \frac{1}{n}\sum_{i=1}^{n} 1 - y_i(w\cdot x_i - b) + \lambda \|w\|^2$$

$$\frac{dJ_i}{dw} = 2\lambda w - y_i \cdot x_i, \quad \frac{dJ_i}{db} = y_i$$

$w = w - \alpha \cdot dw$

$b = b - \alpha \cdot db$

Principal Component Analysis:

Variance - how much variation/spread the data has

$$var(X) = \frac{1}{m}\sum_{i=1}^{n}(X_i - \mu)^2$$

↳ mean normalization

Covariance Matrix - indicates how much two variables vary together

$$cov(X, Y) = \frac{1}{m}\sum_{i=1}^{n}(X_i - \mu)(Y_i - \delta)^T$$

↳ mean of Y_i

$$cov(X, X) = \frac{1}{m}\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^T$$

↳ we want this one

Approach:

1. Mean normalization
2. Calculate $cov(X, X)$
3. Calculate eigenvalues, eigenvectors of $cov(X, X)$
4. Sort eigenvectors according to eigenvalues in decreasing order
5. Choose first k eigenvectors to be new k dimensions
6. Transform (project using dot product) the original n-dimensional points to k dimensions

K-Means Clustering:

Euclidean Distance:

$$d(p,q) = \sqrt{\Sigma(p_i - q_i)^2}$$

Iterative Optimization:

1. Random cluster center (centroid) initialization
2. Repeat until converged {

    Cluster Assignment → Assign points to nearest centroid

    Move Centroid → Set centroid to mean position of each cluster

Naive Bayes Classifier:

Bayes Theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\hookrightarrow P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)} \qquad X = (x_1, x_2 \ldots x_n)$$

Assume all features are mutually independent:

$$P(y|X) = \frac{P(x_1|y) \cdot P(x_2|y) \ldots P(x_n|y)}{P(X)}$$

Select class w/ highest probability:

$$y = \operatorname{argmax}_y P(y|X)$$

$$= \operatorname{argmax}_y \frac{P(x_1 y) \cdot P(x_2|y) \ldots P(x_n|y) \cdot P(y)}{P(X)}$$

$$= \operatorname{argmax}_y \underbrace{P(x_1|y) \cdot P(x_2|y) \ldots P(x_n|y)}_{\text{posterior}} \cdot \underbrace{P(y)}_{\text{prior}}$$

*May cause overflow problems

$$= \operatorname{argmax}_y \log(P(x_1|y)) + \log(P(x_2|y)) + \ldots \log(P(x_n|y)) + \log(P(y))$$

*log changes × to +

Class Conditional Probability $P(x_i|y)$:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

* Gaussian distribution formula