EQUATIONS

A few of the more commonly used equations for Machine Learning/Data Mining

Distance Metrics

Euclidean distance

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Manhattan distance

$$\sum_{i=1}^{n}|x_i - y_i|$$

Hamming distance (x and y are binary vectors)

$$\sum_{i=1}^{n}|x_i - y_i|$$

Minkowski distance

$$\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p}$$

Quadratic equation

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Univariate Statistics

Population mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Standard deviation

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Variance

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

Pre-processing

Normalization

$$z = \frac{x - \mu}{\sigma}$$

Standardizing

$$X_{standarize} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Comparing two vectors

Pearson Correlation

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Spearman Correlation

$$\rho_s = \rho_{X_{[i]},Y_{[i]}} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Cosine Similarity

$$cos(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{x}\cdot\boldsymbol{y}}{||\boldsymbol{x}||\cdot||\boldsymbol{y}||}$$

Co-Variance

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Eigenvector and Eigenvalue

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$$

Binomial distribution

$$p_k = \binom{n}{x}\cdot p^k \cdot (1-p)^{n-k}$$

Gaussian distribution

Univariate

$$p(x) \sim N(\mu|\sigma^2)$$

$$p(x) \sim \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

multivariate

$$p(\boldsymbol{x}) \sim N(\boldsymbol{\mu}|\Sigma)$$

$$p(\boldsymbol{x}) \sim \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

Maximum Likelihood Estimate

Given,

$$D = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$$

Assuming the samples are i.i.d.,

$$p(D|\boldsymbol{\theta}) = p(\boldsymbol{x}_1|\boldsymbol{\theta}) \cdot p(\boldsymbol{x}_2|\boldsymbol{\theta}) \cdot \ \ldots \ p(\boldsymbol{x}_n|\boldsymbol{\theta})$$

$$p(D|\boldsymbol{\theta}) = \prod_{k=1}^{n} p(\boldsymbol{x}_k|\boldsymbol{\theta})$$

The Log likelihood is

$$\Rightarrow l(\theta) = \sum_{k=1}^{n} ln|p(x_k|\theta)$$

Differentiating and solving for (θ)

$$\nabla_{\boldsymbol{\theta}} \equiv \begin{bmatrix} \frac{\partial}{\partial \theta_1} \\ \frac{\partial}{\partial \theta_2} \\ \ldots \\ \frac{\partial}{\partial \theta_p} \end{bmatrix}$$

$$\nabla_{\boldsymbol{\theta}} l(\boldsymbol{\theta}) \equiv \begin{bmatrix} \frac{\partial L(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial L(\boldsymbol{\theta})}{\partial \theta_2} \\ \ldots \\ \frac{\partial L(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix}$$

Linear Discriminant Analysis

In-between class scatter matrix

$$S_w = \sum_{i=1}^{c} S_i$$

where,

$$S_i = \sum_{\boldsymbol{x} \in D_i}^{n} (\boldsymbol{x} - \boldsymbol{m}_i) \ (\boldsymbol{x} - \boldsymbol{m}_i)^T$$

$$\boldsymbol{m}_i = \frac{1}{n_i} \sum_{\boldsymbol{x} \in D_i}^{n} \boldsymbol{x}_k$$

Between class scatter matrix

$$S_b = \sum_{i=1}^{c} (\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T$$

$$\Phi_{lda} = \arg\max_{\Phi} \frac{|\Phi^T S_b \Phi|}{|\Phi^T S_w \Phi|}$$

Multiple Linear Regression

$$\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$$

$$\begin{bmatrix} x_1 & 1 \\ \dots & 1 \\ x_n & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix}$$

$$\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Contour Regression

$$\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$$

$$\begin{bmatrix} x_1 & 1 \\ \dots & 1 \\ x_n & 1 \end{bmatrix} \begin{bmatrix} w \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix}$$

Step 1: Initizalize $\boldsymbol{w}$ using $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Step 2: Repeat until convergence

Step 2a: Reorder $\boldsymbol{X}$ based on latest $\hat{y}$

Step 2b: Estimate $\boldsymbol{w}$ using

$$\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\big(y + \hat{y}_{[i]})$$

Naive Bayes' classifier

Posterior probability:

$$P(\omega_j|x) = \frac{p(x|\omega_j) \cdot P(\omega_j)}{p(x)}$$

$$\Rightarrow \text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

Decision rule:

$$\frac{p(x|\omega_1) \cdot P(\omega_1)}{p(x)} > \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)}$$