Probability Identities

$$P(X,Y) = P(X|Y)P(Y)$$

Product Rule

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Bayes Theorem

$$P(X)=\int_y P(X,y)\,dy=\int_y P(X|y)\,p(y)\,dy$$

Marginalization

$$P(X,Y|Z) = P(X|Y,Z)P(Y|Z) = P(X|Z)P(Y|Z)$$

Conditional Independence

MLE

In statistics, maximum likelihood estimation (MLE) is a method of **estimating the parameters** of a statistical model **given observations**, by finding the **parameter values** that **maximize the likelihood** of **making the observations** given the **parameters**.

$$\boxed{\hat{\theta} = \text{argmax}_{\theta \in \Theta} l(x|\theta)}$$

MAP

In Bayesian statistics, a maximum a posteriori probability (MAP) estimate is an **estimate of an unknown quantity**, that equals the **mode** of the **posterior distribution**.

$$\hat{\theta}_{MAP} = arg\,max\, p(\theta|x) = arg\,max\, p(x|\theta)\, p(\theta)$$

Lagrange

The method of **Lagrange multipliers** is a strategy for finding the **local maxima and minima** of a function subject to **equality constraints**:

maximize f(**x**) subject to g(**x**) = 0

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \lambda g(\boldsymbol{x})$$

[Can also be extended to inequality constraints.]

Newton

The update equation for Newton's method for minimizing a function S w.r.t parameters **β** is

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - \left[\boldsymbol{H}\left(\boldsymbol{\beta}^{(t)}\right)\right]^{-1} \boldsymbol{g}\left(\boldsymbol{\beta}^{(t)}\right)$$

where **g** denotes the gradient vector and **H** the **Hessian** matrix of S.