Contents

Combinations

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k-1)}{k!} = \frac{n!}{(n-k)!k!}$$

Conditional Probability

$P(E \mid F) =$ the probability of E given F has already happened

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Chain Rule

$$P(E \cap F) = P(E \mid F)P(F)$$

Marginalisation

$$P(E) = P(E \cap F_1) + P(E \cap F_2) + \cdots + P(E \cap F_n)$$

$$P(E) = P(E \mid F_1)P(F_1) + P(E \mid F_2)P(F_2) + \cdots + P(E \mid F_n)P(F_n)$$

given

- $F_1, F_2, \ldots, F_n$ are mutually exclusive
- $F_1 \cup F_2 \cup \cdots \cup F_n = S$

Bayes Rule

$$P(E \mid F) = \frac{P(F \mid E)P(E)}{P(F)}$$

Independent Events

$$P(E \cap F) = P(E)P(F)$$

$$P(E \mid F) = P(E)$$

Conditionally Independent

$$P(E \cap F \mid G) = P(E \mid G)P(F \mid G)$$

Binomial Random Variable

Sum of i successes out of n trials.

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, \ldots, n$$

Expected Value of Random Variable

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$$

Variance

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i)$$

with $\mu = E[X]$

Covariance

Say $E[X] = \mu_x$ and $E[Y] = \mu_y$ then

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y]$$

Correlation

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Inequalities

Markov

$$P(X \geq a) \leq \frac{E(X)}{a} \text{ for all } a > 0$$

Chebyshev

$$P(\mid X - \mu \mid \geq k) \leq \frac{\sigma^2}{k^2} \text{ for all } k > 0$$

Chernoff

$$P(X \geq a) \leq \min_{t>0} e^{ta} e^{\log E(e^{tX})}$$

Binomial RV

$$P(X \geq (1 - \delta)np) \leq e^{-np((1+\delta)\log(1+\delta)-\delta)}$$
$$P(X \geq (1 - \delta)\mu) \leq e^{-\mu((1+\delta)\log(1+\delta)-\delta)}$$

Distribution of Sample Mean

$$\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$$

Expected Value

$$E[\bar{X}] = \frac{1}{N}\sum_{k=1}^{N} E[X_k]$$

Variance

$$Var(\bar{X}) = \frac{\sigma^2}{N}$$

Weak Law of Large Numbers

$$P(\mid \bar{X} - \mu \mid \geq \epsilon) \to 0 \text{ as } N \to \infty$$

By Chebyshev's inequality:

$$P(\mid \bar{X} - \mu \mid \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}$$

Continuous Random Variables

Cumulative and Probability Distribution Function

CDF is $F_Y(y)$, PDF is $f_Y(y)$

$$P(a < Y \leq b) = F_Y(b) - F_Y(a)$$
$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt$$
$$P(a < Y \leq b) = \int_{a}^{b} f_Y(t)dt$$

Independent CDF

$$P(X \leq x \wedge Y \leq y) = P(X \leq x)P(Y \leq y)$$

i.e.

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v)du\, dv$$

Conditional PDF

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Chain Rule for PDF

$$f_{XY}(x, y) = f_{X|Y}(x \mid y)f_Y(y) = f_{Y|X}(y \mid x)f_X(x)$$

Marginalisation of PDFs

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy$$

Bayes Rule for PDFs

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y)f_Y(y)}{f_X(x)}$$

Normal Distribution

$Y \sim N(\mu, \sigma^2)$ when it has PDF

$$f_Y(y) = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Central Limit Theorem

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{N}) \text{ as } N \to \infty$$

Linear Regression Model

$$Y = \sum_{i=1}^{m} \Theta^{(i)} x^{(i)} + M$$

where $\vec{\Theta}$ is a vector of unknown (random) parameters and M is random noise

- M is Gaussian with mean 0 and variance 1, $M \sim N(0,1)$
- $\Theta^{(i)}$ is Gaussian with mean 0 and variance λ (where λ is known), $\Theta^{(i)} \sim N(0,\lambda)$

This is equivalent to

$$f_{Y|X,\vec{\Theta}}(y \mid x, \vec{\theta}) = \frac{1}{\sqrt{2\pi}} \exp(-(y - \sum_{i=1}^{m} \theta^{(i)} x^{(i)})^2/2)$$

given $\vec{\Theta} = \vec{\theta}$. Model also assumes $\Theta^{(i)} \sim N(0,\lambda)$ i.e.

$$f_{\Theta^{(i)}}(\theta) \propto \exp(-\theta^2/2\lambda)$$

Parameter Estimation

$$f_{\Theta|D}(\vec{\theta} \mid d) = \frac{f_{D|\Theta}(d \mid \vec{\theta}) f_{\Theta}(\vec{\theta})}{f_D(d)}$$

Maximum Likelihood Estimation

$$\log f_{D|\Theta}(d \mid \vec{\theta}) \propto \log L(\theta) = -\frac{1}{2} \sum_{j=1}^{n} (y_j - \sum_{i=1}^{m} \theta^{(i)} x_j^{(i)})^2$$

$$\theta = \frac{\sum_{j=1}^{n} y_j x_j}{\sum_{j=1}^{n} x_j^2}$$

Maximum a Posteriori (MAP) Estimation

$$\theta = \frac{\sum_{j=1}^{n} y_j x_j}{\frac{1}{\lambda} + \sum_{j=1}^{n} x_j^2}$$

Logistic Regression Model

$$P(Y = 1 \mid \Theta = \vec{\theta}, \vec{X} = \vec{x}) = \frac{1}{1 + \exp(-z)}, z = \sum_{i=1}^{m} \theta^{(i)} x^{(i)}$$

$$P(Y = 0 \mid \Theta = \vec{\theta}, \vec{X} = \vec{x}) = \frac{\exp(-z)}{1 + \exp(-z)}$$

Parameter Estimation

$$P(D = d \mid \Theta = \vec{\theta}) = \prod_{k=1}^{n} (\frac{1}{1 + \exp(-z_k)})^{y_k} (\frac{\exp(-z_k)}{1 + \exp(-z_k)})^{1-y_k}$$

where $z_k = \sum_{i=1}^{m} \theta^{(i)} x_k^{(i)}$

Maximum Likelihood Estimation

$$P(Y = 1 \mid \Theta = \infty, \vec{X} = \vec{x}) = \frac{1}{1 + \exp(-z)}, z = \theta^{(1)} x^{(1)} \begin{cases} 1 & x^{(1)} = -1 \\ 0 & x^{(1)} = 0 \end{cases}$$

$$P(Y = 0 \mid \Theta = \infty, \vec{X} = \vec{x}) = 1 - P(Y = 1 \mid \Theta = \infty, \vec{X} = \vec{x})$$