TECHNO INDIA
SILIGURI INSTITUTE
OF TECHNOLOGY
A Satyam Roychowdhury Initiative

Probability Revisited

Discrete Random Variable, Continuous Random Variable, Joint Probability, Conditional Probability, Probability Distribution

by Dr. Tumpa Banerjee (Dept of MCA)

Discrete Random Variable

- Discrete random variable: the expression $p(A)$ denotes the probability that the event A is true.
- A discrete random variable take any value from a finite or countably infinite set X.
- $p(X = x)$ denotes the probability mass function for the rv X and $\sum p(x) = 1$

Joint Probability

- We define probability of the joint event A and B as follows:
- $p(A, B) = p(A \cap B) = p(A|B)p(B)$
- This is sometimes called product rule.
- Given a joint distribution on two events $p(A, B)$, the marginal distribution is
- $p(A) = \sum_b p(A, B) = \sum_b p(A|B = b)p(B = b)$
- This is called sum rule or rule of total probability.

Joint Probability

- The product rule or chain probability can be written as
- $p(X_{1:N}) = p(X_1)p(X_2|X_1)p(X_3|X_2, X_1) \dots p(X_N|X_1, X_2 \dots X_{N-1})$

Conditional Probability

- The conditional probability for the event A, given that the event B is true
- $p(A|B) = \frac{p(AB)}{p(B)}$ if $p(B) > 0$
- Bayes Rule:
- $p(X = x|Y = y) = \frac{p(X=x,Y=y)}{p(Y=y)} = \frac{p(Y = y|X = x)p(X=x)}{\sum_{\hat{x}} p(Y = y|X = \hat{x})p(X=\hat{x})}$

Example:

- Suppose you are a woman in your 40s, and you decide to have a medical test for breast cancer called a mammogram. If the test is positive, what is the probability you have cancer? Suppose you are told the test has a sensitivity of 80%, which means, if you have cancer, the test will be positive with probability 0.8. the prior probability of having breast cancer is 0.004.

Generative Classifier

- We can generalize the medical diagnosis example to classify feature vectors x of arbitrary type as follows:
- $p(y = c|x, \theta) = \frac{p(x|y = c, \theta) p(y=c|\theta)}{\sum_{\hat{c}} p(x|y = \hat{c}, \theta) p(y=\hat{c}|\theta)}$

Mean and Variance

- The mean or expected value of X is denoted as E(X) or μ
- $E(x) = \sum xp(x)$
- $E(x) = \int xp(x)dx$
- The variance of X is a measure of dispersion, denoted as σ^2
- $E[(X-\mu)^2] = \int (x-\mu)^2 p(x)dx$

Continuous Random Variable

- If X is some uncertain continuous quantity. The probability that X lies in any interval can be written as $P(a \leq X \leq b)$
- We define the cumulative density function of X as $F(a)$
- $P(a \leq X \leq b) = F(b) - F(a)$
- $p(a \leq \mathrm{x} \leq \mathrm{b}) = \int_a^b f(x)dx, where\ f(x)\ is\ pdf\ of\ X$

Discrete Distribution

- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution
- The empirical Distribution

Bernouli Distribution

- A discrete random variable only having two outcome success with probability \$p\$ and failure with probability \$(1-p)\$ follows Bernoulli distribution.
- $P(X = 1) = p$
- And $P(X = 0) = 1 - p$
-

Binomial Distribution

- The binomial distribution is n repetition of the identical experiment, each experiment has only two outcomes: success and failure.
- $P(X = x) = {}^{n}C_{x} p^{x} (1 - p)^{n-x}$

Poisson Distribution

- $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$
- Where $x = 0,1,2,3,4, \ldots$
- And λ is the long-run average

Empirical Distribution

- The empirical distribution for the dataset $D = \{x_1, x_2, \ldots, x_n\}$ of n observation is the function $[0,1] \rightarrow \mathbb{R}$
- $F(x) = \frac{1}{n} \sum \mathbb{I}_n$

Gaussian Distribution

- The widely used distribution in machine learning is the Gaussian distribution. The probability distribution of x is
- $(x|\mu, \sigma^2) = \frac{1}{\sqrt(2\pi\sigma^2)} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$
- Where $\mu = E[x]$ and $\sigma^2 = var(x)$

Some Continuous Distribution

- Gaussian(Normal) Distribution
- Degenerate pdf
- The Laplace Distribution
- The gamma distribution
- Beta Distribution

Covariance and Correlation

- $COV(X,Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right] = E(XY) - E(X)E(Y)$
- $Corr(X,Y) = \frac{COV(X,Y)}{\sqrt{(var(X)Var(Y)}}$

Multivariate Gaussian

- The probability distribution function(pdf) for the multivariate Gaussian in \$D\$ dimension is defined as
- $\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{1/2}} \exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]$