Felix Juefei Xu

Pattern Recognition Theory

Recitation 2: Probability and Statistics

Electrical & Computer ENGINEERING

Topics To Be Covered

- Basic Probability Theory
  - Elementary Stuff
  - Bayes Rule
- Random Variables (RVs)
  - PDFs and CDFs
  - Mean and Variance
  - Commonly Used PDFs
- Joint Distributions (>1 RV)
- Conditional Probability Revisited
- MATLAB functions

The Basic Stuff

- Define probability of an event as *P*(*A*)

$$\boxed{P(A) = \lim_{n \to \infty} \frac{n_A}{n}}$$

- Axioms of probability
  - 0<=*P*(*A*)<=1
  - *P* (Certain Event) = 1, *P* (Impossible Event) =0
  - If *A* and *B* are Mutually Exclusive i.e.

$$\boxed{P[A \cap B] = 0 \text{ then } P[A \cup B] = P[A] + P[B]}$$

- *A* and *B* are Independent Events if *P*(*AB*) = *P*(*A*)P(*B*)

Conditional Probability

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

Bayes Rule

$$P(B) = P(B|A_1)P(A_1) + ...P(B|A_n)P(A_n)$$

Total Probability Rule

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + ...P(B|A_n)P(A_n)}$$

Bayes Rule + Total Probability Rule

Random Variable Preliminaries

- An RV represents the probability of different events and hence takes on different values with probabilities that sum up to 1
- An RV can be Continuous, Discrete or Mixed
- Cumulative Distribution Function (CDF) – Non Decreasing Function

$$F_X(x) = P(X \leq x)$$

$$F(+\infty) = 1, F(-\infty) = 0$$

$$F(x_2) - F(x_1) = P(x_2 < x \leq x_1)$$

- Probability Density (Mass) Function (PDF or PMF)

$$f_X(x) = \frac{d}{dx}(F_X(x))$$

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

$$F_X(x) = \sum_{x \leq x_i} f_X(x_i)$$

Mean and Variance

- Mean is also known as expected value or expectation

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Continuous RV

$$\sum_{-\infty}^{+\infty} x f_X(x)$$

Discrete RV

- Variance is second moment about mean

$$\sigma^2 = E[(X - E(X)^2] = E(X^2) - E^2(X)$$

$$\int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)$$

Continuous RV

$$\sum_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)$$

Discrete RV

Properties of Mean and Variance

- Expectation is a linear operator
- $E(X + c) = E(X) + E(c) = E(X) + c$
- $E(cX) = cE(X)$
- $E(X + Y) = E(X) + E(Y)$
- $E(XY) = E(X)E(Y)$ only if X and Y are uncorrelated or independent
- $\mathrm{var}(aX) = a^2\mathrm{var}(X)$

Discrete Densities

Bernoulli

$$\boxed{f_X(x) = x^p(1-x)^{(1-p)} \quad X = 0,1}$$

Binomial

$$\boxed{P(X=k) = \binom{n}{k} p^k q^{n-k} \quad p+q=1 \quad k=0,1,2\ldots n}$$

Geometric

$$\boxed{P(X=k) = pq^{k-1} \quad k=1,2,3\ldots\infty}$$

Poisson

$$\boxed{P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad k=0,1,2\ldots\infty}$$

Continuous Densities

Uniform

$$f_X(x) = \frac{1}{b-a} \quad a \le x \le b$$
$$= 0 \quad otherwise$$

Normal

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty \le x \le +\infty$$

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy \triangleq G(\frac{x-\mu}{\sigma})$$

$$G(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

Exponential

$$f_X(x) = \lambda e^{-\lambda x} \quad x \ge 0$$
$$= 0 \quad otherwise$$

Joint Distributions – Bivariate

$$\boxed{F_{X,Y}(x,y) = P(X \leq x, Y \leq y)}$$

CDF

$$\boxed{f_{X,Y}(x,y) = \frac{\delta^2 F_{X,Y}(x,y)}{\delta x \delta y}}$$

PDF

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

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

Marginal PDFs

Joint Distributions – Bivariate

$$\boxed{cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)}$$

Covariance

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

Correlation Coefficient

$$\boxed{E(X,Y) = E(X)E(Y)}$$

Uncorrelated

$$\boxed{f_{X,Y}(x,y) = f_X(x) f_Y(y)}$$

Independent

Joint Distributions – Multivariate

$$F_{\underline{X}}(\underline{x}) = F_{X_1,X_2...X_n}(x_1,x_2...x_n) = P(X_1 \le x_1 ... X_n \le x_n) = \int_{-\infty}^{x_n} ... \int_{-\infty}^{x_1} f_{X_1,X_2...X_n}(x_1,x_2...x_n)dx_1...dx_n$$

$$F_{\underline{X}}(-\infty \ ... \ -\infty) = 0 \quad F_{\underline{X}}(\infty \ ... \ \infty) = 1$$

CDF

$$f_{\underline{X}}(\underline{x}) = P(X_1 = x_1 ... X_n = x_n) = f_{X_1,X_2...X_n}(x_1,x_2...x_n) = \frac{dF_{\underline{X}}(\underline{x})}{d\underline{X}} = \frac{\delta^n F_{X_1,X_2...X_n}(x_1,x_2...x_n)}{\delta x_1 ... \delta x_n}$$

$$f_{\underline{X}}(\underline{x}) \ge 0$$

$$\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} f_{X_1,X_2...X_n}(x_1,x_2...x_n)dx_1...dx_n = 1$$

PDF

Joint Distributions – Multivariate

$$\boxed{\underline{\mu} = E[\underline{X}] = E[X_1 \; \ldots \; X_n]^T = [E[X_1] \; \ldots \; E[X_n]]^T}$$

Mean Vector

$$\underline{\Sigma} = \begin{bmatrix} cov(X_1,X_1) & cov(X_1,X_2) & \ldots & cov(X_1,X_n) \\ cov(X_2,X_1) & cov(X_2,X_2) & \ldots & cov(X_2,X_n) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ cov(X_n,X_1) & cov(X_n,X_2) & \ldots & cov(X_n,X_n) \end{bmatrix} = \begin{bmatrix} \sigma^2_{X_1} & cov(X_1,X_2) & \ldots & cov(X_1,X_n) \\ cov(X_2,X_1) & \sigma^2_{X_2} & \ldots & cov(X_2,X_n) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ cov(X_n,X_1) & cov(X_n,X_2) & \ldots & \sigma^2_{X_n} \end{bmatrix}$$

$$\boxed{\underline{\Sigma} = E[(\underline{X} - E(\underline{X}))(\underline{X} - E(\underline{X}))^T] = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T] = E[\underline{X}\underline{X}^T] - \underline{\mu}\underline{\mu}^T}$$

Covariance Matrix

$$\boxed{f_{X_1 \ldots X_n}(x_1 \ldots x_n) = \prod_i^n (f_{X_i}(x_i))}$$

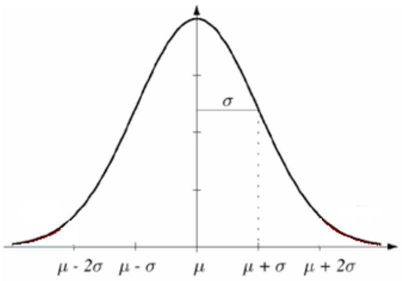
Independence

Electrical & Computer ENGINEERING

Gaussian (Normal) Distribution

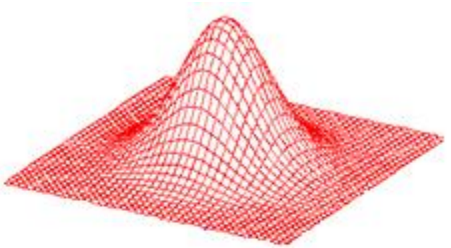
Univariate Normal Distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$



Multivariate Normal Distribution

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\underline{\mathbf{\Sigma}}|^{1/2}} exp[\frac{-(\underline{x}-\underline{\mu})^T \underline{\mathbf{\Sigma}}^{-1}(\underline{x}-\underline{\mu})}{2}]$$



$$\text{If } \underline{X} \text{ is } N(\underline{\mu}, \underline{\mathbf{\Sigma}}) \text{ then } \underline{Y} = \underline{\mathbf{A}X} \text{ is } N(\underline{\mathbf{A}}\underline{\mu}, \underline{\mathbf{A\Sigma A}}^T)$$

Conditional Probability Revisited

$$f_{X|Y}(x|y) = P(X = x|Y = y) = f_{X,Y}(x, y)/f_Y(y) = P(X = x, Y = y)/P(Y = y)$$

Bayes Rule

$$f(y|x) = f(x, y)/f(x)$$

$$f(x|y) = f(x, y)/f(y)$$

$$f(x, y) = f(x|y)f(y) = f(y|x)f(x)$$

Simplified Notation

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{f(x|y)f(y)}{f(y)} = \frac{f(y|x)f(x)}{f(y)} = \frac{f(y|x)f(x)}{\int_{-\infty}^{\infty} f(y|x)f(x)dx}$$

The Grand Scheme

Useful MATLAB Functions

- rand/randn
- randperm
- pdf
- normpdf
- mvnpdf
- cdf
- erf, erfc, erfinv, erfcinv

References

- Useful Denitions and Results in Probability Theory - Notes By Prof. Vijaykumar Bhagavatula for Pattern Recognition
- Athanasios Papoulis, S. Unnikrishna Pillai, “Probability, Random Variables and Stochastic Processes,” TMH 4th edition, 2002
- Richard O. Duda, Peter E. Hart, David G. Stork, “Pattern Classification,” Wiley 2nd edition, 2007
- MATLAB Help