ECE368: Probabilistic Reasoning

Aman Bhargava

January-April 2021

Contents

0.1 Introduction and Course Information

Course Information

- Professors: Prof. Saeideh Parsaei Fard and Prof. Foad Sohrabi
- Course: Engineering Science, Machine Intelligence Option
- Term: 2021 Winter

Main Course Topics

- Vector, temporal, and spatial models.
- Classification and regression model training.
- Bayesian statistics, frequentist statistics.

Chapter 1

Review Topics

See ECE286 notes for further reference

1.1 Review of Probability Functions

Probability Mass Function: For *discrete random variables*, $P_X(x)$denotes the probability that random variable X takes on value x.

Probability Density Function: For *continuous random variables*, the probability $\Pr\{X \in [x_1, x_2]\}$ is given by $\int_{x_1}^{x_2} f_X(x)dx$.
Joint PMF's and PDF's are similarly defined.

Marginal Probability Distributions: Given joing PMF $P_{X,Y}(x,y)$ or PDF $f_{X,Y}(x,y)$, we can **marginalize** them as follows:

$$P_X(x) = \sum_{y \in Y} P_{X,Y}(x,y) \tag{1.1}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \tag{1.2}$$

Conditional Probability Functions:

$$P_{Y|X}(y,x) = \frac{P_{X,Y}(x,y)}{P_X(x)} \tag{1.3}$$

Prior Probability: Probability **before** an additional observation is made (hence *prior*). Example: $P_X(x)$.

Posterior Probability: Probability **after** an observation is made (hence *post*erior). Example: $P_{X|Y}(x, y)$.

Bayes Rule:

$$P(B|A) = P(A|B)\frac{P(B)}{P(A)} \tag{1.4}$$

1.2 Expectation, Correlation, and Independence

Expectation Value: $\mathbb{E}[x] = \sum_{x \in X} P_X(x) = \int_{-\infty}^{\infty} x f_X(x) dx$

Law or Large Numbers: $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} x_i = \mathbb{E}[X]$

Variance:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[x])^2] \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned} \tag{1.5}$$

Covariance:

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}_{XY}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned} \tag{1.6}$$

Correlation Coefficient:

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \tag{1.7}$$

- $\rho_{XY} \in [-1, 1]$
- $\rho > 0$ indicates positive correlation (line of best fit has positive slope).
- $\rho < 0$ indicates negative correlation.
- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ **iff** X, Y **are uncorrelated**.

Independence

Theorem 1 *Independence Random variables X, Y are independent **iff***

$$P_{XY}(x, y) = P_X(x) \cdot P_Y(y) \tag{1.8}$$

This also means that $\rho_{XY} = 0$, $P(X|Y) = P(X)$, etc.

1.3 Laws of Large Numbers

Weak Law: Sample mean converges to the mean.

Strong Law: If $\{x_i\}$ are **independent, identically distributed** (i.i.d.) random variables with mean μ, then the **probability of** the sample mean $= \mu$ is 1 as $n \to \infty$.

Chapter 2

Parameter Estimation

2.1 Estimation Terminology