Machine Learning - Assignment I

Review of Probability and Statistics

Souhaib BEN TAIEB

24 February 2020

Please provide justifications for every step you take in your derivations.

Question 1

Consider a sample space Ω comprising four possible outcomes:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}.$$

Consider the three events E, F and G defined as follows:

$$E = \{\omega_1\}, \quad F = \{\omega_1, \omega_2\}, \quad G = \{\omega_1, \omega_2, \omega_3\}.$$

Suppose their probabilities are

$$P(E) = \frac{1}{10} \quad P(F) = \frac{5}{10} \quad P(G) = \frac{7}{10}$$

Now, consider a fourth event H defined as follows;

$$H = \{\omega_2, \omega_4\}$$

Find $P(H)$.

Question 2

Consider a sample space Ω comprising three possible outcomes:

$$\Omega = \{\omega_1, \omega_2, \omega_3\}.$$

Suppose the three possible outcomes are assigned the following probabilities:

$$P(\omega_1) = \frac{1}{5} \quad P(\omega_2) = \frac{2}{5} \quad P(\omega_3) = \frac{2}{5}$$

Define the events

$$E = \{\omega_1, \omega_2\}, \quad F = \{\omega_1, \omega_3\},$$

and denote by E^c the complement of E.

Compute $P(F|E^c)$, the conditional probability of F given E^c.

Question 3

An economics consulting firm has created a model to predict recessions. The model predicts a recession with probability 80% when a recession is indeed coming and with probability 10% when no recession is coming. The unconditional probability of falling into a recession is 20%. If the model predicts a recession, what is the probability that a recession will indeed come?

Question 4

(a) Let X be a discrete random variable taking values in $R_X = \{0, 1, 2, 3, 4\}$ with uniform distribution. Compute $P(1 \leq X < 4)$.

(b) Let X be a continuous random variable taking values in $R_X = [0, 1]$. Let its probability density function $f_X(x)$ be

$$f_X(x) = \begin{cases} 2x & \text{if } x \in R_X \\ 0 & \text{if } x \notin R_X \end{cases}$$

Compute $P(\frac{1}{4} \leq X < \frac{1}{2})$.

Question 5

(a) Let X be a continous random variable with uniform distribution on the interval $[1, 3]$. Compute the expected value and variance of X.

(b) Let X be a discrete random variable taking values in $R_X = \{1, 2, 3\}$. Let its probability mass function $p_X(x)$ be

$$p_X(x) = \begin{cases} x/6 & \text{if } x \in R_X \\ 0 & \text{if } x \notin R_X \end{cases}$$

Compute the expected value and variance of X.

Question 6

Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Let $Z = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Show that

- $E[Z] = \mu$
- $\mathrm{Var}(Z) = \frac{\sigma^2}{n}$

Question 7

- Let $\mu = 1$, and consider the following four scenarios: (a) $n = 10, \sigma = 1$, (b) $n = 10, \sigma = 5$, (c) $n = 1000, \sigma = 1$ and (d) $n = 1000, \sigma = 5$.
- For each scenario, repeat the following procedure $10,000$ times:
  - Generate n i.i.d. realizations $X_1, X_2, \ldots, X_n$ where $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ for $i = 1, 2, \ldots, n$. You can assume the X_i variables are normally distributed.
  - Compute $Z = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- For each scenario, plot a histogram of the $10,000$ values for Z.
- For each scenario, compute the (empirical) mean and variance of Z, and discuss the results as a function of n and σ using your answer to Question 3.

Question 8

Suppose that $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \text{Geom}(p)$, i.e. the samples have a geometric distribution with parameter p. A geometric distribution is the distribution of the number of coin flips needed to see one head.

- Write down the likelihood as a function of the observed data $X_1, X_2, \ldots, X_n$, and the unknown parameter p.
- Compute the MLE of p. In order to do this you need to find a zero of the derivative of the likelihood, and also check that the second derivative of the likelihood at the point is negative.

TURN IN

- Your `.Rmd` file (which should knit without errors and without assuming any packages have been pre-loaded)
- Your pdf file that results from knitting the Rmd.
- DUE: March 8, 11:55pm (late submissions not allowed), loaded into Moodle