Homework 1 - Introduction to Machine Learning for Engineers

kipngeno koech - bkoech

January 19, 2025

1 Probability

1. Suppose W is a Gaussian random variable with distribution $N(\mu, \sigma^2)$ and U a uniform random variable over the interval $[a, b]$. Assuming that W and U are independent, what is the expected value $\mathbb{E}[Z]$ and variance $\mathrm{Var}[Z]$ of $Z = 3W + 2U$?

   - **Expected value:**
     $$\mathbb{E}[Z] = \mathbb{E}[3W + 2U] = 3\mathbb{E}[W] + 2\mathbb{E}[U]$$
     Since W is Gaussian with mean μ and U is uniform over $[a, b]$ with mean $\frac{a+b}{2}$:
     $$\mathbb{E}[Z] = 3\mu + 2\left(\frac{a+b}{2}\right) = \mathbf{3\mu + (a + b)}$$
   - **Variance:** if a random variable X, is scaled by a constant a, then the variance of the scaled random variable is a^2 times the variance of the original random variable. Therefore:
     $$\mathrm{Var}[Z] = \mathrm{Var}[3W + 2U] = 3^2\mathrm{Var}[W] + 2^2\mathrm{Var}[U]$$
     Since W is Gaussian with variance σ^2 and U is uniform over $[a, b]$ with variance $\frac{(b-a)^2}{12}$:
     $$\mathrm{Var}[Z] = 9\sigma^2 + 4\left(\frac{(b-a)^2}{12}\right) = 9\sigma^2 + \frac{(b-a)^2}{3}$$

2. Consider the following joint distribution between the random variable X, which takes values T or F, and the random variable Y, which takes values a, b, c, or d.

$P(X, Y)$	$Y = a$	$Y = b$	$Y = c$	$Y = d$
$X = T$	0.1	0.2	0.1	0.1
$X = F$	0.1	0.1	0.2	0.1

   (a) **Marginal distribution P_Y:**
   $$\Pr(Y = a) = \Pr(X = T, Y = a) + \Pr(X = F, Y = a) = 0.1 + 0.1 = \mathbf{0.2}$$
   $$\Pr(Y = b) = \Pr(X = T, Y = b) + \Pr(X = F, Y = b) = 0.2 + 0.1 = \mathbf{0.3}$$
   $$\Pr(Y = c) = \Pr(X = T, Y = c) + \Pr(X = F, Y = c) = 0.1 + 0.2 = \mathbf{0.3}$$
   $$\Pr(Y = d) = \Pr(X = T, Y = d) + \Pr(X = F, Y = d) = 0.1 + 0.1 = \mathbf{0.2}$$

   (b) **Conditional probability $\Pr(X = T \mid Y \in \{b, c, d\})$:** since we are conditioning on Y being in the set $\{b, c, d\}$, we need to find the probability of $X = T$ and Y being in the set $\{b, c, d\}$ and divide it by the probability of Y being in the set $\{b, c, d\}$:
   this is the bayes theorem:
   $$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(B)}$$
   $$\Pr(Y \in \{b, c, d\}) = \Pr(Y = b) + \Pr(Y = c) + \Pr(Y = d) = 0.3 + 0.3 + 0.2 = 0.8$$
   $$\Pr(X = T \cap Y \in \{b, c, d\}) = \Pr(X = T, Y = b) + \Pr(X = T, Y = c) + \Pr(X = T, Y = d) = 0.2 + 0.1 + 0.1 = 0.4$$
   $$\Pr(X = T \mid Y \in \{b, c, d\}) = \frac{\Pr(X = T \cap Y \in \{b, c, d\})}{\Pr(Y \in \{b, c, d\})} = \frac{0.4}{0.8} = \mathbf{0.5}$$

2 Linear Algebra

1. Let $A_k \in \mathbb{R}^{n \times n}$ for $k = 1, \ldots, K$ such that $A_k = A_k^\top$, i.e., each A_k is a symmetric, n-dimensional square matrix. Suppose all A_k have the exact same set of eigenvectors $u_1, u_2, \ldots, u_n$ with the corresponding eigenvalues $\alpha_{k1}, \ldots, \alpha_{kn}$ for each A_k. Write down the eigenvectors and their corresponding eigenvalues for the following matrices:

   (a) $C = \sum_{k=1}^{K} A_k$

   $$\text{Eigenvectors: } \mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}$$

   $$\text{Eigenvalues: } \sum_{\mathbf{k=1}}^{\mathbf{K}} \alpha_{\mathbf{k1}}, \sum_{\mathbf{k=1}}^{\mathbf{K}} \alpha_{\mathbf{k2}}, \ldots, \sum_{\mathbf{k=1}}^{\mathbf{K}} \alpha_{\mathbf{kn}}$$

   (b) $D = A_i^{-1} A_j A_i$, where $i \neq j$ and $i, j \in \{1, 2, \ldots, K\}$. Here we assume A_i is invertible. A matrix A is similar to a matrix B if there exists an invertible matrix P such that $A = P^{-1}BP$. Similar matrices have the same eigenvalues. Therefore, the eigenvalues of D are the same as the eigenvalues of A_j.
   Since A_j has the same eigenvectors as A_i, the eigenvectors of D are the same as the eigenvectors of A_i.

   $$\text{Eigenvectors: } \mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}$$

   For the eigenvalues:

   $$\text{Eigenvalues: } \alpha_{\mathbf{j1}}, \alpha_{\mathbf{j2}}, \ldots, \alpha_{\mathbf{jn}}$$

2. Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ be given, and $\text{col}(A)$ be the column space of A. For a given value of m, under what conditions on b, $\text{col}(A)$, and $\text{rank}(A)$ will the equation $Ax = b$ have:
   The column space of a matrix A is the set of all possible linear combinations of the columns of A.:

   $$\text{col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

   The rank of a matrix is the dimension of the column space of the matrix.

   (a) **No solution:**

   The equation $Ax = b$ has no solution if $b \notin \text{col}(A)$. This means that b is not a linear combination of the columns of A.

   (b) **Exactly one solution:**

   The equation $Ax = b$ has exactly one solution if $b \in \text{col}(A)$ and $\text{rank}(A) = n$. This means that b is a linear combination of the columns of A

   (c) **Infinitely many solutions:**

   The equation $Ax = b$ has infinitely many solutions if $b \in \text{col}(A)$ and $\text{rank}(A) < n$.