HW2

January 27, 2024

1 Assignment 1 (Due Tues 2/6). May be submitted online by midnight through canvas (pdf of hand written notes or latex are preferred), hand calculations may alternatively be submitted during Tues lecture.

1.1 Problem 1 [10 Points]

We fill extend the binary classification code presented in class 1/25 (available on the course repo under Lecture 2.ipynb) to implement a multi-class regression problem. For this you will find **S**2.5 of Murphy to be helpful.

For the Iris dataset there are 3 classes (Virginica, Versicolor, Setosa). We would like to train a model that will reproduce the plot in figure 2.13.

1.1.1 Subproblem 1 [7 Points]

From the panda dataframe presented in class, extract the petal width and length features, and define them as input labels X_1 and X_2, respectively. Construct a tensorflow model which assigns a distribution to the categorical random variable

$$Y = \begin{cases} 0, & \text{if class } = \text{Virginica} \\ 1, & \text{if class } = \text{Versicolor} \\ 2, & \text{if class } = \text{Setosa} \end{cases}$$

which we will model via

$$p(Y|\mathbf{x}, \mathbf{W}, \mathbf{b}) = Cat\left(Y|\mathcal{S}, \mathbf{W}\mathbf{x} + \mathbf{b}\right)$$

where $\mathcal{S}$ is a softmax, $\mathbf{W} \in \mathbb{R}^{4\times 2}$, $\mathbf{x} \in \mathbb{R}^2$, and $\mathbf{b} \in \mathbb{R}^4$. Use the log-sum-exp trick to ensure your model is stable, using the example code from class together with $\mathcal{S}$2.5.4 of Murphy to understand the syntax and extension from binary to categorical.

1.1.2 Subproblem 2 [3 Points]

Visualize the decision boundary of your dataset (i.e. reproduce fig. 2.13) by generating a contour plot of the function $f(x_1, x_2) = \underset{x_1, x_2}{\text{argmax}} p(Y|\mathbf{x}, \mathbf{W}, \mathbf{b})$ over the range of petal widths and lengths provided in the dataset. Superimpose over the contour plot a scatter plot of the input data, colored by their class.

1.2 Problem 2 [10 Points]

Consider a coin that comes up heads with probability p and tails with probability $1-p$. Let q_n be the probability of obtaining an even number of heads in n independent tosses.

1.2.1 Subproblem 1 [5 Points]

Derive a recursion that relates q_n to q_{n-1} and establish the formula (4 pts)

$$q_n = \frac{1+(1-2p)^n}{2}$$

To hunt for a recurrence relation, you will want to use conditioning to relate q_n to q_{n-1}. After obtaining an expression $q_n = f_n(q_{n-1})$ you can then attempt to solve the recurrence relation (1 pt).

1.2.2 Subproblem 2 [5 Points]

Let X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} Ce^{-(ax+by)} & x, y \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

where, $a, b > 0$ are constants

- Determine the constant C
- Find the marginal density of X and Y and use them to demonstrate whether X and Y are independent.
- Find $\mathbb{E}(Y \mid X > \frac{\exp(a^2+b^2)}{a^4+b^4})$

1.3 Problem 2 [5 points] - Working with distributions

Let $S_1, S_2, \ldots, S_n$ be a partition of the sample space Ω, i.e. $\Omega = \cup_i S_i$ and $S_i \cap S_j = \emptyset$. - Show that for any event A,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap S_i)$$

- Use the previous part to show that, for events A, B and C,

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) + \mathbb{P}(A \cap B^c \cap C^c) - \mathbb{P}(A \cap B \cap C)$$

- Prove that for any two events A and B, we have

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$$

[You will need to mathematically prove these facts using the list of set operations we covered in the probability crash course. Drawing a diagram is not enough.]

1.4 Problem 3 [10 Points] - Manipulating distribution functions

1.4.1 Subproblem 1 [5 Points]

- If $X_1, X_2, \ldots, X_n$ are independent random variables having the same probability density function $f_X(x)$, what is the probability density function for the random variable $Y = \min\{X_1, X_2, \ldots, X_n\}$?
- Consider two continuous random variables Y and Z and a random variable X that is equal to Y with a probability p and equals Z with a probability $1-p$. Obtain the pdf of X interms of the pdf's of Y and Z.

1.4.2 Subproblem 2 [5 Points]

The Laplace distribution is given by

$$p(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

Consider a mixture of three Laplace distributions: $p(x) = \alpha p_1(x) + \beta p_2(x) + \gamma p_3(x)$

where $\alpha, \beta, \gamma \in [0,1]$ are mixture weights satisfying $\alpha + \beta + \gamma = 1$ and $p_1(x)$, $p_2(x)$ and $p_3(x)$ are Laplace distributions with different parameters $(\mu_1, b_1) \neq (\mu_2, b_2) \neq (\mu_3, b_3)$.

Derive the expectation and variance of $p(x)$, analytically, using their definitions.

1.5 Problem 4 [10 Points] - Computing with Normal distributions

As mentioned in class the Gaussian has nice properties which makes it a fundamental tool in statistical infererence. The standard normal $\mathcal{N}(\mathbf{x}; \mu, \mathbf{\Sigma})$ is defined as

$$\mathcal{N}(\mathbf{x}; \mu, \mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1} (\mathbf{x}-\mu)\right]$$

Recall that a Gaussian distribution is uniquely specified by its expectation and covariance.

- Prove that if $x \in \mathbb{R}^d$ is normally distributed, every affine transformation $y = Ax + b$ also has a Gaussian distribution. Find its mean and covariance.
- Analytically find the KL divergence $\mathbb{KL}(P||Q)$ between two multivariate normal distributions $p(x) \sim \mathcal{N}(\mathbf{x}; \mu_1, \Sigma_1)$ and $q(x) \sim \mathcal{N}(\mathbf{x}; \mu_2, \mathbf{\Sigma}_2)$.