Homework #0

CSEP 590B: Explainable AI
Prof. Su-In Lee
Due: 4/4/22 11:59 PM

1 Survey questions (5 points)

To help us get to know you better, please answer the following questions about your experience with machine learning (ML) and explainable AI (XAI).

(a) [1 points] What is your level of ML experience (e.g., new, proficient, expert)? What ML-related courses have you taken, during your current degree or otherwise?

(b) [1 points] Where did you get experience with ML (e.g., school, personal study, work)? If you use ML for your job, what domain do you focus on (e.g., medicine, finance, advertising)?

(c) [1 points] What ML models do you use most often (e.g., linear models, random forests, neural networks)? Are there any ML models you would like to learn more about?

(d) [1 points] Describe a past ML project that you found most fascinating, either because of the impact, the effectiveness (or ineffectiveness) of the method, or the elegance of the tools involved.

(e) [1 points] What XAI tools are you familiar with, if any? What problems do you hope to solve with XAI?

2 Probability review (10 points)

(a) [1 points] Consider a continuous random variable $\mathbf{x} \in \mathbb{R}$ with probability density function $p(\mathbf{x})$. Define the random variable's expected value, $\mathbb{E}[\mathbf{x}]$.

(b) [1 points] Define the random variable's variance, $\text{Var}(\mathbf{x})$.

(c) [1 points] Given n independent and identically distributed (i.i.d.) samples $x_1, \ldots, x_n \sim p(\mathbf{x})$, write an estimator for the expected value $\mathbb{E}[\mathbf{x}]$.

(d) [1 points] Show that the estimator from (c) is an *unbiased* estimator (i.e., its expected value is equal to $\mathbb{E}[\mathbf{x}]$, the value being estimated). **Hint:** recall the linearity of expectations property (see here).

(e) [2 points] Show how the estimator's variance from (c) compares to $\text{Var}(\mathbf{x})$. **Hint:** recall the properties for the variance of a sum of random variables (see here).

(f) [2 points] What does the (weak) law of large numbers (LLN) say about the estimator from (c)?

(g) [2 points] Given a second random variable $\mathbf{y}$, how do we determine the conditional probability $p(\mathbf{y} \mid \mathbf{x})$? What does it mean when we have $p(\mathbf{y}) = p(\mathbf{y} \mid \mathbf{x})$?

3 Calculus review (7 points)

(a) [1 points] Consider a continuous function $f(t)$, or $f : \mathbb{R} \mapsto \mathbb{R}$. Write the definition of the derivative $\frac{df}{dt}(t)$.

(b) [1 points] Describe the geometric interpretation of the derivative.

(c) [1 points] Consider a continuous function $g(x, y, z)$, or $g : \mathbb{R}^3 \mapsto \mathbb{R}$. Write the definition of the partial derivative $\frac{\partial g}{\partial y}(x, y, z)$.

(d) [2 points] For the function $g(x, y, z)$ and an input value $(x_0, y_0, z_0) \in \mathbb{R}^3$, find the "direction of greatest increase" and describe what this means. How does this relate to gradient descent?

(e) [2 points] For a function $h(x, y, z)$ defined as $h(x, y, z) = f(g(x, y, z))$, use the chain rule to find $\frac{\partial h}{\partial y}$ in terms of f and g. What is the relevance of chain rule for training neural networks?

4 ML model review (8 points)

(a) [2 points] For a dataset with n examples denoted $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^n$, derive the solution for a linear regression model fit with the standard least squares loss (assuming no intercept term, for simplicity). If we instead fit a logistic regression model with discrete labels $Y \in \{0, 1\}^n$ and log-loss, how does the model fitting procedure differ?

(b) [2 points] What are the differences between decision trees, random forests, and gradient-boosted trees (e.g., XGBoost)?

(c) [2 points] Write the equations to produce predictions for a multi-layer perceptron (MLP) with input $x \in \mathbb{R}^d$, one hidden layer of size p, sigmoid activations, and K output probabilities (for K possible classes). Indicate the size of each learnable parameter. **Hint:** rather than writing a single equation, write a sequence of equations where each output provides the input for the next operation.

(d) [1 points] List the sequence of layers used in the VGG-16 neural network architecture. What is the shape of the last convolutional layer's output? (**Hint:** you can find this in the original paper, or you can find the architecture details in a blog post.)

(e) [1 points] For a neural network with parameters $\theta \in \mathbb{R}^L$, denote the prediction given an input x as $f(x; \theta)$ and the loss for a single prediction as $\ell(f(x; \theta), y)$. Write the update step that we use to improve the parameters θ repeatedly over the course of training.