CSCI 5521: Machine Learning Fundamentals (Spring 2024)

Quiz 1 (Thursday, Feb 8)

Due on Gradescope at 02:00 PM, Friday, Feb 9

Instructions:

- This quiz has 3 questions, 30 points, on 1 page.
- Please **write your name & ID on your submission pages**.

1. (**6 points**) Supervised learning has a wide range of applications. For example, with the popularity of wearable devices, such as FitBit and Apple Watch, one of their core functionalities is to identify whether a time period contains physical activity or rest based on **heart rate data** collected from the device. Please model this problem as a two-class classification task and answer the following questions.

    (a) Name two relevant features to this two-class classification task (i.e., whether a time period contains physical activity or rest). Any reasonable features are acceptable.

    (b) What are the labels in this two-class classification task?

2. (**10 points**) What could we do to reduce overfitting in a polynomial regression model? Select all the option(s) that apply.

    (a) Collect 30% more new training data.

    (b) Sample half of the original training data and **only** use that as the new training data.

    (c) Double the size of the original test data (by collecting new data) and replace as the new test data.

    (d) Increase polynomial degree.

    (e) Decrease polynomial degree so that the model becomes a linear regression model.

3. (**14 points**) The conditional probability density functions of two classes C_1 and C_2 are shown in the figure below, with $P(x|C_1) \sim \mathcal{N}(1.5, 0.8)$ and $P(x|C_2) \sim \mathcal{N}(0.9, 1.3)$.

    (a) Assuming the priors are equal, predict which class (C_1 or C_2) the data point $x = 0.5$ (illustrated with the green dot) belongs to. Briefly explain why.

    (b) What if the priors are $P(C_1) = 0.6$ and $P(C_2) = 0.4$, respectively?