Name+Email ID Jon-Michael Hoang | hoang339

CSCI 5521: Machine Learning Fundamentals (Spring 2022)

Quiz 1 (Thursday, Feb 10)

Due on Gradescope at 02:00 PM, Friday, Feb 11

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, one can define a robot (vs non-robot) based on a number of traits. Please model the question as a two-class classification task and answer the following questions.

   (a) Name two relevant features to this two-class classification task. (Any reasonable features are acceptable).

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

a) $x = \begin{bmatrix} \text{instructions received} \\ \text{actions performed in real world} \end{bmatrix}$ b) $r = \begin{cases} 1, & \text{if } x \text{ is positive (is robot)} \\ 0, & \text{if } x \text{ is negative (not robot)} \end{cases}$

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

   IDK

   overfitting?

   (a) Decrease polynomial degree.

   (b) Change to a more complex model (*e.g.,* a model with more parameters).

   (c) Add new training data and keep test data the same.

   (d) Add new test data and keep training data the same. underfitting

   (e) Sample half of the original training data as new training data.

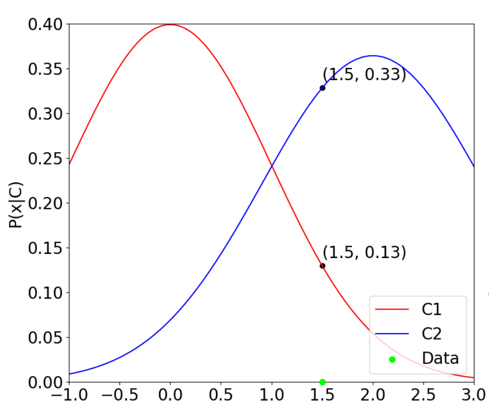
   same generalization

   wouldn't it result in the same thing?

3. (**14 points**) The conditional probability density functions of two classes C_1 and C_2 are shown in the figure below, with $P(x|C_1) \sim \mathcal{N}(0, 1)$ and $P(x|C_2) \sim \mathcal{N}(2, 1.2)$.

   (a) Assuming the priors are equal, predict which class (C_1 or C_2) the data point $x = 1.5$ (illustrated with the green dot) belongs to. Briefly explain why.

   (b) What if the priors are $P(C_1) = 0.9$ and $P(C_2) = 0.1$, respectively? (Note: High-level explanation is good, but you can use formulations if it helps explain).



a) $P(C_1) = 1.5$, $P(C_2) = 1.5$, $x = 1.5$

$P(x|C_1) = .13$, $P(x|C_2) = .33$

$P(C_1|x) = \frac{1.5 * .13}{1.5(.13) + 1.5(.33)} = \boxed{0.283}$

$P(C_2|x) = \frac{1.5 * .33}{1.5(.13) + 1.5(.33)} = \boxed{0.717}$

∴ $x = 1.5$ would most likely belong to C_2.

b) If the priors are weighted s.t. $P(C_1) = .9$ and $P(C_2) = .1$,

$$P(C_1 \mid x) = \frac{.9 * .13}{.9(.13) + .1(.33)} = \boxed{.78}$$

$$P(C_2 \mid x) = \frac{.1 * .33}{.9(.13) + .1(.33)} = \boxed{.22}$$

$\boxed{\therefore \text{ it would belong to } C_1}$