80629A Homework #1
Laurent Charlin
Due date: October 25, 2019

Instructions:
• Please include your name and HEC ID with submission.
• The homework is due by 4pm on the due date. Please hand in a paper version of your homework and leave it in room 4.632 (CSC).
• The homework is worth 20% of the course’s final grade.
• Assignments are to be done individually.

1 General ML (9%)
For a–b, consider a continuous smooth joint function over features and class labels such that no example is ever duplicated exactly. For c–e, the answers should fit within 3–4 lines.

a. (2pt) For a fixed model complexity, sketch a graph of the typical behaviour of training error rate versus training set size in a learning system. Add to this graph a curve showing the typical behaviour of test error rate (on an iid infinite test set) versus training set size, on the same axes. Mark a horizontal line showing the Bayes error. Indicate on your vertical axes where zero error is and draw your graphs with increasing error upwards and increasing training set size rightwards.

b. (2pt) For a fixed range of model complexity (from very simple to very complex), sketch a graph of training set size versus the model complexity which achieves the best test performance (on an iid infinite test set).

c. (1pt) Explain why it is sensible to make the test set big when you are evaluating the generalization performance of a learning procedure.

d. (2pt) Explain what is the difference between a parameter and a hyper-parameter.

e. (2pt) Explain the role of the prior in a probabilistic model.

2 Classification (11%)

a. (2pt) What is “naive” about the naive Bayes model?

b. Using scikit-learn, fit the multinomialNB and the GaussianNB models on the Wine classification dataset. Details regarding the data are available here https://scikit-learn.org/stable/datasets/index.html#wine-dataset. Please use the default parameters for all.

• (2pt) What is the main conceptual difference between both models? What is the right model to use for the Wine dataset (please provide a short justification)?

• (3pt) [Bonus] Derive the maximum likelihood update for the parameters of class conditional distribution in a multinomialNB model. (You can use the BernoulliNB instead for 1 pt.)

• (2pt) Report the train and test errors that you obtain for each model (i.e., you need to report 4 values, 2 for each model). To split your data use the following scikit-learn function call (it will create a test set that’s 20% of your original data):

    train_test_split(x, y, shuffle=True, test_size=0.2, random_state=1234)
• (2pt) Which is the most useful feature? Please justify your answer as there may be more than one correct answer.

c. (3pt) Fit a neural network to the Wine dataset (same as above). You are now allowed to tune the default hyper-parameters of the neural network to obtain better generalization. Report the train/test error you obtained and the hyper-parameters you chose (reuse the same train and test splits as above). Which is the most useful (data) feature? Briefly justify your answer.