ELL784/EEL709: Introduction to Machine Learning

Assignments Quiz

Please answer on this question paper itself; do not use a separate answer book

Maximum marks: 12, Form: $\boxed{\text{A}}$

Name: _____

Student Number: _____

TA: _____

Date: _____

Section 1. Multiple choice questions

Each question may have any number of correct choices, including zero. Encircle all choices you believe to be correct (1 mark for each correct choice, -0.5 for each incorrect choice). No justification is required.

1. We have seen two error functions that are used for neural networks: sum-of-squares error (SSE), and cross-entropy error (CEE). Suppose we are training a neural network for binary classification. Which of the follwing are true?

   (a) SSE cannot be used; it works only for regression.

   (b) CEE should be preferred to SSE, because CEE is closer to classification error, which is what we really care about.

   (c) CEE should be preferred to SSE, because CEE also takes into account the magnitude of error, rather than just right/wrong.

   (d) Both CEE and SSE can give good results, but in principle CEE might be slightly preferable because it corresponds to maximising the likelihood of the data.

2. When doing multi-variable regression (multiple inputs, single output), which of the following might be reasonable ways of doing feature selection, assuming you are confident the features are largely independent of each other?

   (a) Try eliminating features one-at-a-time; pick the features that lead to the greatest increase in training error upon elimination.

   (b) Normalise each feature to have mean 0 and variance 1; then just pick the features with the highest positive weights in the full model.

   (c) Try eliminating features one-at-a-time; pick the features that lead to the greatest increase in testing error upon elimination.

   (d) Pick the features with the highest absolute weights in the full model trained on the raw inputs.

   (e) Try eliminating features one-at-a-time; pick the features that lead to the greatest increase in cross-validation error upon elimination.

3. You have seen how both a neural network and PCA can be used to map data to a lower-dimensional representation. Suppose, for your labeled data set, you contstruct two representations: one consisting of the top M principal components, and one via training a neural network with a single hidden layer with M units. Now, you use both these M-dimensional representations as input features to train an SVM to classify your data. Which of the following are true?

   (a) Generally, the neural network representation should work better.

   (b) Generally, the PCA representation should work better.

   (c) There is no general or systematic reason why one representation should work better than the other.

   (d) The parameter settings for the SVM will generally determine which one works better.

4. You are doing regularised polynomial regression, and you find that your current model has similarly high error during both training and testing. Which of the following might generally be expected to reduce error?

    (a)   Since this is underfitting, I can try decreasing the number of training data points to get a better fit.

    (b)   Since this is overfitting, I can try increasing the number of training data points to get a better fit.

    (c)   Increasing the order of the polynomial.

    (d)   Using less noisy training data.

    (e)   Weakening the regularisation.

Section 2. Numerical questions

5. Consider the following confusion matrix for a three-class data set on which an *unsupervised* clustering has been done:

	Actual class 1	Actual class 2	Actual class 3
Placed in cluster 1	31	34	2
Placed in cluster 2	19	3	34
Placed in cluster 3	1	19	17
Placed in cluster 4	9	43	12

As you did in the assignment, evaluate the overall accuracy of this clustering by identifying each cluster with the most frequent label in it.   [2]

6. Consider the following results from a soft-margin RBF SVM with parameters C and σ:

C	σ	Training error	Cross-validation error	C	σ	Training error	Cross-validation error
1	5	0.25	0.26	10	5	0.17	0.24
1	10	0.28	0.29	10	10	0.21	0.21
5	5	0.21	0.22	50	5	0.06	0.37
5	10	0.23	0.24	50	10	0.12	0.32

(a) Give the values of C and σ for the best-fit case.   [1]

(b) In the above table, how many cases each of overfitting and underfitting are there, relative to the best-fit case?   [1]

(c) Suppose you are at an underfit case. Should you increase or decrease C, in order to try and improve the fit?   [1]

(d) In terms of the bias-variance trade-off, which of the given parameter settings corresponds to the highest bias?   [1]