ELL409: Machine Intelligence and Learning

Assignments Quiz, Form: A

Maximum marks: 12

(Answer all questions on this question paper. Read all section-specific instructions carefully.)

Name: _____

Entry Number: _____

Section 1. Multiple choice questions

Instructions: Each question may have any number of correct choices. Clearly mark (tick) all choices you believe to be correct (1.5 marks for each correct choice, −0.5 for each incorrect choice).

1. In training an SVM with a polynomial kernel, one typically has two hyperparameters: the order of the polynomial d, and the slack penalty C. Suppose one does a grid search on these and obtains a contour plot showing pairs of values which correspond to the same cross-validation error. Moving along a given contour in the direction of increasing d, C will generally be

   (a) Increasing
   (b) Decreasing
   (c) Increasing for the part of the contour corresponding to overfitting, decreasing for the part corresponding to underfitting
   (d) Decreasing for the part of the contour corresponding to overfitting, increasing for the part corresponding to underfitting

2. A neural network with 30 hidden units gave a cross-validation accuracy of 96% on a classification data set; when the number of units was increased to 40, the cross-validation accuracy was recorded as 91%. Which of the following are likely to increase the generalisation performance of the 40-hidden-unit network in this case?

   (a) Making the backpropagation convergence criterion more stringent
   (b) Decreasing the learning rate η
   (c) Adding a second hidden layer of similar dimension to the current one
   (d) Early stopping of backpropagation

3. Can cross-validation error be regarded as a reasonable estimate of testing error?

   (a) Yes, even if the error has been obtained after hyperparameter tuning
   (b) Yes, but only when no hyperparameter tuning has been done
   (c) Yes, but only when the number of folds is not too large
   (d) No, never

4. Suppose you've been given a regression data set generated from a polynomial function plus some unknown kind of noise. You fit a regression function to it by minimising sum-of-squares error. In which of the following circumstances will the resulting model be expected to accurately recover the underlying polynomial, assuming you have provided for sufficient data and model complexity?

   (a) Only when the noise is Gaussian
   (b) Only when the noise is symmetric about zero
   (c) Only when the noise is zero-mean
   (d) Always

5. Based on previous patients, a hospital created a data set with symptoms as features and tagged them with just 2 different class labels: *healthy* and *diseased*. The hospital created a train-test split and requested ELL409 students to come up with a machine learning algorithm for this classification task. The hospital will select the algorithm which can achieve the best performance (as per some suitable metric) on the test set. Assuming that the data is highly imbalanced (most patients belong to the *healthy* class), which of the following performance metrics (individually) would be suitable for the hospital to use?

   (a) Recall ($TP/(TP+FN)$) for the *diseased* class
   (b) Precision ($TP/(TP+FP)$) for the *diseased* class
   (c) F1-score (harmonic mean of recall and precision) for the *diseased* class
   (d) Binary classification accuracy over the two classes

6. Suppose you try quadratic (L2) regularisation on a regression model fit to data sets of different sizes sampled from the same population (*e.g.*, the data sets of size 20 and 100 you used in Assignment 1). The error function used is sum-of-squares error, $E(\mathbf{w}) = \sum_{n=1}^{N}(y(\mathbf{x}_n) - t_n)^2$. For each data set, you tune the regularisation parameter λ using cross-validation. What is the expected relation between the value of λ obtained and N, the size of the training data set used?

   (a) λ should increase as N increases
   (b) λ should decrease as N increases
   (c) The relation between λ and N depends on the dimension of the parameter vector $\mathbf{w}$
   (d) In general, λ should not depend on N

7. Suppose you wish to normalise/rescale your features (either to the range $[0, 1]$, or via z-scoring) before training a classifier on them. Which of the following approaches would be suitable to assess the validation performance of the trained classifier?

   (a) Perform the feature normalisation once on your entire data set, then assess the classifier via cross-validation on this set.
   (b) Do cross-validation on the entire data set, but for each fold, separately normalise the features for the training and validation sets.
   (c) Do cross-validation on the entire data set, and for each fold, compute the normalisation parameters for each feature (min and max, or μ and σ) only on the respective training set; then apply these same parameters to rescale/normalise the features for the validation set.
   (d) Split the data set into just a single training and validation set each; normalise the features separately for both sets, then train and validate as usual.

Section 2. Short-answer question

8. With reference to the earlier question, the hospital selected the model supplied by student S. The model outperformed every other student's model in the class on the test set; however, it turned but to perform very badly when deployed at the hospital. S used the following strategy: they started with a model M_1 trained on the training set and evaluated it on the test set using the performance metric identified in the earlier question. S then improved the model (to get M_2) after carefully analysing the misclassified instances on the test set and adjusting the model accordingly. After K such rounds of improvement S was able to get state-of-the-art performance with model M_K on the test set and supplied the model M_K to the hospital.

   Can you figure out what might have gone wrong? **[1.5]**

ELL409: Machine Intelligence and Learning

Assignments Quiz, Form: B

Maximum marks: 12

(Answer all questions on this question paper. Read all section-specific instructions carefully.)

Name: _____

Entry Number: _____

Section 1. Multiple choice questions

Instructions: Each question may have any number of correct choices. Clearly mark (tick) all choices you believe to be correct (1.5 marks for each correct choice, −0.5 for each incorrect choice).

1. Suppose you wish to normalise/rescale your features (either to the range $[0, 1]$, or via z-scoring) before training a classifier on them. Which of the following approaches would be suitable to assess the validation performance of the trained classifier?

   (a) Perform the feature normalisation once on your entire data set, then assess the classifier via cross-validation on this set.

   (b) Do cross-validation on the entire data set, but for each fold, separately normalise the features for the training and validation sets.

   (c) Do cross-validation on the entire data set, and for each fold, compute the normalisation parameters for each feature (min and max, or μ and σ) only on the respective training set; then apply these same parameters to rescale/normalise the features for the validation set.

   (d) Split the data set into just a single training and validation set each; normalise the features separately for both sets, then train and validate as usual.

2. In training an SVM with a polynomial kernel, one typically has two hyperparameters: the order of the polynomial d, and the slack penalty C. Suppose one does a grid search on these and obtains a contour plot showing pairs of values which correspond to the same cross-validation error. Moving along a given contour in the direction of increasing d, C will generally be

   (a) Increasing

   (b) Decreasing

   (c) Increasing for the part of the contour corresponding to overfitting, decreasing for the part corresponding to underfitting

   (d) Decreasing for the part of the contour corresponding to overfitting, increasing for the part corresponding to underfitting

3. Suppose you try quadratic (L2) regularisation on a regression model fit to data sets of different sizes sampled from the same population (*e.g.*, the data sets of size 20 and 100 you used in Assignment 1). The error function used is sum-of-squares error, $E(\mathbf{w}) = \sum_{n=1}^{N}(y(\mathbf{x}_n) - t_n)^2$. For each data set, you tune the regularisation parameter λ using cross-validation. What is the expected relation between the value of λ obtained and N, the size of the training data set used?

   (a) λ should increase as N increases

   (b) λ should decrease as N increases

   (c) The relation between λ and N depends on the dimension of the parameter vector $\mathbf{w}$

   (d) In general, λ should not depend on N

4. A neural network with 30 hidden units gave a cross-validation accuracy of 96% on a classification data set; when the number of units was increased to 40, the cross-validation accuracy was recorded as 91%. Which of the following are likely to increase the generalisation performance of the 40-hidden-unit network in this case?

   (a) Making the backpropagation convergence criterion more stringent

   (b) Decreasing the learning rate η

   (c) Adding a second hidden layer of similar dimension to the current one

   (d) Early stopping of backpropagation

5. Suppose you've been given a regression data set generated from a polynomial function plus some unknown kind of noise. You fit a regression function to it by minimising sum-of-squares error. In which of the following circumstances will the resulting model be expected to accurately recover the underlying polynomial, assuming you have provided for sufficient data and model complexity?

   (a) Only when the noise is Gaussian
   (b) Only when the noise is symmetric about zero
   (c) Only when the noise is zero-mean
   (d) Always

6. Can cross-validation error be regarded as a reasonable estimate of testing error?

   (a) Yes, even if the error has been obtained after hyperparameter tuning
   (b) Yes, but only when no hyperparameter tuning has been done
   (c) Yes, but only when the number of folds is not too large
   (d) No, never

7. Based on previous patients, a hospital created a data set with symptoms as features and tagged them with just 2 different class labels: *healthy* and *diseased*. The hospital created a train-test split and requested ELL409 students to come up with a machine learning algorithm for this classification task. The hospital will select the algorithm which can achieve the best performance (as per some suitable metric) on the test set. Assuming that the data is highly imbalanced (most patients belong to the *healthy* class), which of the following performance metrics (individually) would be suitable for the hospital to use?

   (a) Recall ($TP/(TP+FN)$) for the *diseased* class
   (b) Precision ($TP/(TP+FP)$) for the *diseased* class
   (c) F1-score (harmonic mean of recall and precision) for the *diseased* class
   (d) Binary classification accuracy over the two classes

Section 2. Short-answer question

8. With reference to the earlier question, the hospital selected the model supplied by student S. The model outperformed every other student's model in the class on the test set; however, it turned but to perform very badly when deployed at the hospital. S used the following strategy: they started with a model M_1 trained on the training set and evaluated it on the test set using the performance metric identified in the earlier question. S then improved the model (to get M_2) after carefully analysing the misclassified instances on the test set and adjusting the model accordingly. After K such rounds of improvement S was able to get state-of-the-art performance with model M_K on the test set and supplied the model M_K to the hospital.

   Can you figure out what might have gone wrong? **[1.5]**

ELL409: Machine Intelligence and Learning

Assignments Quiz, Form: C

Maximum marks: 12

(Answer all questions on this question paper. Read all section-specific instructions carefully.)

Name: _____

Entry Number: _____

Section 1. Multiple choice questions

Instructions: Each question may have any number of correct choices. Clearly mark (tick) all choices you believe to be correct (1.5 marks for each correct choice, −0.5 for each incorrect choice).

1. Suppose you've been given a regression data set generated from a polynomial function plus some unknown kind of noise. You fit a regression function to it by minimising sum-of-squares error. In which of the following circumstances will the resulting model be expected to accurately recover the underlying polynomial, assuming you have provided for sufficient data and model complexity?

   (a) Only when the noise is Gaussian
   (b) Only when the noise is symmetric about zero
   (c) Only when the noise is zero-mean
   (d) Always

2. Suppose you try quadratic (L2) regularisation on a regression model fit to data sets of different sizes sampled from the same population (*e.g.*, the data sets of size 20 and 100 you used in Assignment 1). The error function used is sum-of-squares error, $E(\mathbf{w}) = \sum_{n=1}^{N}(y(\mathbf{x}_n) - t_n)^2$. For each data set, you tune the regularisation parameter λ using cross-validation. What is the expected relation between the value of λ obtained and N, the size of the training data set used?

   (a) λ should increase as N increases
   (b) λ should decrease as N increases
   (c) The relation between λ and N depends on the dimension of the parameter vector $\mathbf{w}$
   (d) In general, λ should not depend on N

3. Based on previous patients, a hospital created a data set with symptoms as features and tagged them with just 2 different class labels: *healthy* and *diseased*. The hospital created a train-test split and requested ELL409 students to come up with a machine learning algorithm for this classification task. The hospital will select the algorithm which can achieve the best performance (as per some suitable metric) on the test set. Assuming that the data is highly imbalanced (most patients belong to the *healthy* class), which of the following performance metrics (individually) would be suitable for the hospital to use?

   (a) Recall ($TP/(TP + FN)$) for the *diseased* class
   (b) Precision ($TP/(TP + FP)$) for the *diseased* class
   (c) F1-score (harmonic mean of recall and precision) for the *diseased* class
   (d) Binary classification accuracy over the two classes

4. A neural network with 30 hidden units gave a cross-validation accuracy of 96% on a classification data set; when the number of units was increased to 40, the cross-validation accuracy was recorded as 91%. Which of the following are likely to increase the generalisation performance of the 40-hidden-unit network in this case?

   (a) Making the backpropagation convergence criterion more stringent
   (b) Decreasing the learning rate η
   (c) Adding a second hidden layer of similar dimension to the current one
   (d) Early stopping of backpropagation

5. In training an SVM with a polynomial kernel, one typically has two hyperparameters: the order of the polynomial d, and the slack penalty C. Suppose one does a grid search on these and obtains a contour plot showing pairs of values which correspond to the same cross-validation error. Moving along a given contour in the direction of increasing d, C will generally be

   (a) Increasing
   (b) Decreasing
   (c) Increasing for the part of the contour corresponding to overfitting, decreasing for the part corresponding to underfitting
   (d) Decreasing for the part of the contour corresponding to overfitting, increasing for the part corresponding to underfitting

6. Suppose you wish to normalise/rescale your features (either to the range $[0, 1]$, or via z-scoring) before training a classifier on them. Which of the following approaches would be suitable to assess the validation performance of the trained classifier?

   (a) Perform the feature normalisation once on your entire data set, then assess the classifier via cross-validation on this set.
   (b) Do cross-validation on the entire data set, but for each fold, separately normalise the features for the training and validation sets.
   (c) Do cross-validation on the entire data set, and for each fold, compute the normalisation parameters for each feature (min and max, or μ and σ) only on the respective training set; then apply these same parameters to rescale/normalise the features for the validation set.
   (d) Split the data set into just a single training and validation set each; normalise the features separately for both sets, then train and validate as usual.

7. Can cross-validation error be regarded as a reasonable estimate of testing error?

   (a) Yes, even if the error has been obtained after hyperparameter tuning
   (b) Yes, but only when no hyperparameter tuning has been done
   (c) Yes, but only when the number of folds is not too large
   (d) No, never

Section 2. Short-answer question

8. With reference to the earlier question, the hospital selected the model supplied by student S. The model outperformed every other student's model in the class on the test set; however, it turned but to perform very badly when deployed at the hospital. S used the following strategy: they started with a model M_1 trained on the training set and evaluated it on the test set using the performance metric identified in the earlier question. S then improved the model (to get M_2) after carefully analysing the misclassified instances on the test set and adjusting the model accordingly. After K such rounds of improvement S was able to get state-of-the-art performance with model M_K on the test set and supplied the model M_K to the hospital.

   Can you figure out what might have gone wrong? **[1.5]**

ELL409: Machine Intelligence and Learning

Assignments Quiz, Form: D

Maximum marks: 12

(Answer all questions on this question paper. Read all section-specific instructions carefully.)

Name: _____

Entry Number: _____

Section 1. Multiple choice questions

Instructions: Each question may have any number of correct choices. Clearly mark (tick) all choices you believe to be correct (1.5 marks for each correct choice, −0.5 for each incorrect choice).

1. Suppose you've been given a regression data set generated from a polynomial function plus some unknown kind of noise. You fit a regression function to it by minimising sum-of-squares error. In which of the following circumstances will the resulting model be expected to accurately recover the underlying polynomial, assuming you have provided for sufficient data and model complexity?

   (a) Only when the noise is Gaussian
   (b) Only when the noise is symmetric about zero
   (c) Only when the noise is zero-mean
   (d) Always

2. Can cross-validation error be regarded as a reasonable estimate of testing error?

   (a) Yes, even if the error has been obtained after hyperparameter tuning
   (b) Yes, but only when no hyperparameter tuning has been done
   (c) Yes, but only when the number of folds is not too large
   (d) No, never

3. A neural network with 30 hidden units gave a cross-validation accuracy of 96% on a classification data set; when the number of units was increased to 40, the cross-validation accuracy was recorded as 91%. Which of the following are likely to increase the generalisation performance of the 40-hidden-unit network in this case?

   (a) Making the backpropagation convergence criterion more stringent
   (b) Decreasing the learning rate η
   (c) Adding a second hidden layer of similar dimension to the current one
   (d) Early stopping of backpropagation

4. In training an SVM with a polynomial kernel, one typically has two hyperparameters: the order of the polynomial d, and the slack penalty C. Suppose one does a grid search on these and obtains a contour plot showing pairs of values which correspond to the same cross-validation error. Moving along a given contour in the direction of increasing d, C will generally be

   (a) Increasing
   (b) Decreasing
   (c) Increasing for the part of the contour corresponding to overfitting, decreasing for the part corresponding to underfitting
   (d) Decreasing for the part of the contour corresponding to overfitting, increasing for the part corresponding to underfitting

5. Based on previous patients, a hospital created a data set with symptoms as features and tagged them with just 2 different class labels: *healthy* and *diseased*. The hospital created a train-test split and requested ELL409 students to come up with a machine learning algorithm for this classification task. The hospital will select the algorithm which can achieve the best performance (as per some suitable metric) on the test set. Assuming that the data is highly imbalanced (most patients belong to the *healthy* class), which of the following performance metrics (individually) would be suitable for the hospital to use?

   (a) Recall ($TP/(TP+FN)$) for the *diseased* class
   (b) Precision ($TP/(TP+FP)$) for the *diseased* class
   (c) F1-score (harmonic mean of recall and precision) for the *diseased* class
   (d) Binary classification accuracy over the two classes

6. Suppose you try quadratic (L2) regularisation on a regression model fit to data sets of different sizes sampled from the same population (*e.g.*, the data sets of size 20 and 100 you used in Assignment 1). The error function used is sum-of-squares error, $E(\mathbf{w}) = \sum_{n=1}^{N}(y(\mathbf{x}_n) - t_n)^2$. For each data set, you tune the regularisation parameter λ using cross-validation. What is the expected relation between the value of λ obtained and N, the size of the training data set used?

   (a) λ should increase as N increases
   (b) λ should decrease as N increases
   (c) The relation between λ and N depends on the dimension of the parameter vector $\mathbf{w}$
   (d) In general, λ should not depend on N

7. Suppose you wish to normalise/rescale your features (either to the range $[0, 1]$, or via z-scoring) before training a classifier on them. Which of the following approaches would be suitable to assess the validation performance of the trained classifier?

   (a) Perform the feature normalisation once on your entire data set, then assess the classifier via cross-validation on this set.
   (b) Do cross-validation on the entire data set, but for each fold, separately normalise the features for the training and validation sets.
   (c) Do cross-validation on the entire data set, and for each fold, compute the normalisation parameters for each feature (min and max, or μ and σ) only on the respective training set; then apply these same parameters to rescale/normalise the features for the validation set.
   (d) Split the data set into just a single training and validation set each; normalise the features separately for both sets, then train and validate as usual.

Section 2. Short-answer question

8. With reference to the earlier question, the hospital selected the model supplied by student S. The model outperformed every other student's model in the class on the test set; however, it turned but to perform very badly when deployed at the hospital. S used the following strategy: they started with a model M_1 trained on the training set and evaluated it on the test set using the performance metric identified in the earlier question. S then improved the model (to get M_2) after carefully analysing the misclassified instances on the test set and adjusting the model accordingly. After K such rounds of improvement S was able to get state-of-the-art performance with model M_K on the test set and supplied the model M_K to the hospital.

   Can you figure out what might have gone wrong? **[1.5]**