Model training, Model evaluation (I), and Model evaluation (II) - Performance measures.

Total points 39/42

Email *

shinoda.c.i@gmail.com

✓ Which of the following best defines model training in machine learning? * 1/1

- a) The process of testing a trained model on new, unseen data
- b) The process of feeding data into a machine learning algorithm to adjust the model's parameters ✓
- c) The process of visualizing the results of a machine learning model's predictions

Feedback

b) This is typically done by minimizing the difference between the model's predicted output and the actual output

✓ Which of the following is a potential problem when using low-degree polynomial functions in machine learning? *1/1

- The model may be too complex
- The model may be too simple ✓
- The model may require too much computational power to train

✗ If the model's predictions are far off, what might you do to make them better? *0/1

- [x] a) Retrain the model using a larger and more diverse dataset. ✓
- [ ] b) You can't fix a model whose predictions are far off.
- [x] c) Try a different training approach. For example, if you used a supervised approach, try an unsupervised approach. ✗
- [x] d) Retrain the model, but use only the features you believe have the strongest predictive power for the label. ✓

Correct answer

- [x] a) Retrain the model using a larger and more diverse dataset.
- [x] d) Retrain the model, but use only the features you believe have the strongest predictive power for the label.

Feedback

b) Incorrect. It is usually possible to fix a model whose predictions are off. Most models require multiple rounds of training until they make useful predictions.

c) Incorrect. A different training approach would not produce better predictions.

✓ Model generalization is the ability to perform well on new (previously unobserved) inputs. *1/1

- (•) True ✓
- ( ) False

✓ The train set must be used to train the model, i.e., adjust the parameters, while the test set must be used to verify the model's performance. *1/1

- (●) True ✓
- ( ) False

✓ Smaller training error indicates better generalization. * 1/1

- ( ) True
- (●) False ✓

Feedback

Smaller test error indicates better generalization.

✓ It is better to evaluate the model using part of the training data. * 1/1

- ( ) True
- (●) False ✓

Feedback

To measure generalization, it is important to evaluate the model using data that has not been used during the model training process.

✓ Which of the following may cause overfitting? * 1/1

- [ ] The model capacity is too small
- [x] The model capacity is too large ✓
- [x] The training set includes a large amount of dirty data ✓
- [x] The number of training sets is small ✓

✗ High training error reduces the accuracy of the model and produces underfitting. How to improve the model fit in this case? *0/1

- [x] Increase the amount of data ✗
- [x] Feature Engineering (E.g., add better features) ✓
- [x] Reduce regularization parameters ✓
- [ ] Reduce the model capacity

Correct answer

- [x] Feature Engineering (E.g., add better features)
- [x] Reduce regularization parameters

✓ The generalization error will continue to decrease as the complexity of the model increases *1/1

- ○ True
- ◉ False ✓

Feedback

Correct. Increasing the model complexity too much may cause overfitting, which causes the generalization error to increase.

✓ The training error will continue to decrease as the model complexity increases *1/1

- ◉ True ✓
- ○ False

Feedback

Correct. The higher the model capacity, the easier it is for the training algorithm to fit it to the training data - Notice, however, that a good fit to the training data does not imply good generalization.

✓ The test error will keep getting smaller as the complexity of the model increases. *1/1

- ○ True
- ◉ False ✓

Feedback

Correct. Increasing the model complexity too much may cause overfitting, which causes the test error to increase.

✓ Regularization is a technique that can be used to prevent overfitting in machine learning models. *1/1

- ◉ True ✓
- ○ False

✓ Overfitting is more likely when you have a huge amount of data to train. * 1/1

- ○ True
- ◉ False ✓

✓ Your model is overfitting the training dataset. How to improve the model generalization in this case? *1/1

- [x] Increase the amount of training data ✓
- [ ] Reduce regularization parameters
- [x] Reduce the model capacity ✓
- [x] Perform data cleaning to reduce the noise in the training data ✓

✓ If a model has a large bias on the test set and a small variance, it means that the model? *1/1

- ○ Is overfitting
- ○ May be overfitting and may be underfitting
- ○ Is a good fit for the dataset
- ◉ Is underfitting ✓

✓ About the important considerations when designing and deploying machine learning models, choose the true alternatives *1/1

- [x] Generalization refers to the ability of a model to perform well on new, unseen data ✓
- [x] Interpretability refers to the ability of a model to provide explanations or insights into how its predictions are generated ✓
- [x] Prediction speed refers to the time it takes for a model to make predictions once it has been trained ✓
- [x] Practicability refers to the overall feasibility and usability of a machine learning model in a practical setting ✓

✓ In machine learning, the _____ set is typically used to train a model, while the _____ set is used to evaluate how the model would perform on new, unseen data. *1/1

- validation, test
- test, validation
- ◉ training, test ✓
- test, training

✓ What can happen if the degree of a polynomial used to fit a machine learning model is too high? *1/1

- The training error might be high and the test error low
- ◉ The training error might be low while the test error might be high ✓
- Both the training error and the test error will be high

✓ Suppose you are trying to train a linear model to solve a task in which the target variable has a quadratic relationship with the input variables. Which of the following is most likely to happen? *1/1

- ○ The model will perfectly fit the data
- ◉ The model will underfit the data and perform poorly on both the training and test sets ✓
- ○ The model will overfit the data and perform well on the training set, but poorly on the test set
- ○ The model will be unable to learn anything from the data and produce random predictions

Feedback

Indeed. A linear model is not expressive enough to capture the quadratic relationship between the input features and the target

✓ Regarding overfitting in machine learning: * 1/1

- ○ It happens when the model is too simple and underperforms on the training data
- ◉ It happens when the model is too complex and starts to fit the noise in the training data ✓
- ○ It happens when the model performs well on both the training and test data
- ○ It happens when the model is unable to learn from the training data

✓ Which methods can be used to address both underfitting and overfitting in a polynomial model? *1/1

- [ ] Decreasing the amount of data used for training
- [x] Adjusting the degree of the polynomial function ✓
- [ ] Increasing the complexity of the model

✓ How does increasing a model's capacity affect its ability to fit the training data and generalize to new data? *1/1

- ( ) Increasing capacity might lead to underfitting.
- (•) Increasing capacity can lead to overfitting, causing the model to fit the training data well but show poor generalization performance ✓
- ( ) Increasing capacity has no effect on a model's ability to fit the training data or generalize to new data

Feedback

Indeed. When a model has high capacity, it has more parameters and can represent complex functions. This means that it has more flexibility to fit the training data more closely. However, this can lead to overfitting,

✓ Which of the following is an advantage of using a larger training dataset? * 1/1

- (●) It reduces overfitting and improves generalization performance ✓
- ( ) It increases model capacity and reduces underfitting
- ( ) It reduces training time and computation cost

Feedback

Indeed. Increasing the size of the training dataset can reduce overfitting by providing the model with more diverse examples that better represent the distribution of the data

✓ What is the goal of regularization in machine learning? * 1/1

- (●) To reduce overfitting and improve generalization performance ✓
- ( ) To increase model capacity and reduce underfitting
- ( ) To reduce the size of the training dataset

Feedback

Correct. Regularization methods add constraints to the model's parameters during training, which helps to prevent overfitting by reducing the variance of the model

✓ Which of the following indicators are used to evaluate regression models? *1/1

- Recall rate
- Confusion matrix
- ◉ Mean square error ✓
- Accuracy

✓ Which of the following is NOT a true statement about RMSE, MAE, MSE, and R2? *1/1

- RMSE and MAE both measure the difference between predicted and actual values.
- MSE is calculated as the average of squared differences between predicted and actual values
- R2 ranges from −∞ to 1, with a higher value indicating a better fit
- ◉ MAE is always smaller than RMSE ✓

Feedback

Correct. In fact, there is no guarantee that MAE will always be smaller than RMSE

✓ Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable? *1/1

- AUC-ROC
- Accuracy
- Logloss
- ◉ Mean-Squared-Error ✓

✓ Mean squared error is a metric typically used to evaluate the performance of classification models. *1/1

- True
- ◉ False ✓

Feedback

Mean squared error is used to evaluate the performance of regression models.

✓ Consider the following confusion matrix for the next 3 exercises: * 1/1

What is the accuracy of the model based on the confusion matrix?

	Predicted Dog	Predicted Cat	Predicted Bird
Actual Dog	150	20	10
Actual Cat	30	140	30
Actual Bird	20	10	160

- (●) 78.95% ✓
- ( ) 83.33%
- ( ) 81.67%
- ( ) 85.00%
- ( ) 88.33%

Feedback

Indeed, accuracy = # correct predictions / total number of predictions

Hence, accuracy = (150+140+160) / (570) = 78.95%

✓ What is the precision of the cat class based on the confusion matrix? * 1/1

- 78.00%
- 82.35% ✓
- 84.00%
- 88.00%

Feedback

Indeed, the precision of the cat class is equal to 140 / (20+140+10) = 82.35%

✓ What is the recall of the bird class based on the confusion matrix? * 1/1

- 84.21% ✓
- 86.96%
- 90.48%
- 92.31%

Feedback

Indeed, the recall for the bird class = 160 / (20+10+160) = 84.21%

✗ In which of the following scenarios would a high accuracy value suggest that the ML model is doing a good job? *0/1

- (●) a) An expensive robotic chicken crosses a very busy road a thousand times per day. An ML model evaluates traffic patterns and predicts when this chicken can safely cross the street with an accuracy of 99.99%. ✗
- ( ) b) A deadly, but curable, medical condition afflicts .01% of the population. An ML model uses symptoms as features and predicts this affliction with an accuracy of 99.99%.
- ( ) c) In the game of roulette, a ball is dropped on a spinning wheel and eventually lands in one of 38 slots. Using visual features (the spin of the ball, the position of the wheel when the ball was dropped, the height of the ball over the wheel), an ML model can predict the slot that the ball will land in with an accuracy of 4%

Correct answer

- (●) c) In the game of roulette, a ball is dropped on a spinning wheel and eventually lands in one of 38 slots. Using visual features (the spin of the ball, the position of the wheel when the ball was dropped, the height of the ball over the wheel), an ML model can predict the slot that the ball will land in with an accuracy of 4%

Feedback

a) Incorrect. A 99.99% accuracy value on a very busy road strongly suggests that the ML model is far better than chance. In some settings, however, the cost of making even a small number of mistakes is still too high. 99.99% accuracy means that the expensive chicken will need to be replaced, on average, every 10 days. (The chicken might also cause extensive damage to cars that it hits.)

b) Incorrect. Accuracy is a poor metric here. After all, even a "dumb" model that always predicts "not sick" would still be 99.99% accurate. Mistakenly predicting "not sick" for a person who actually is sick could be deadly.

✓ Consider a classification model that separates email into two categories: "spam" or "not spam". This model outputs a continuous spam probability value and decides on the class based on a classification threshold. If you raise the classification threshold, what will happen to precision? *1/1

- a) Probably decrease.
- b) Probably increase. ✓
- c) Definitely decrease.
- d) Definitely increase.

Feedback

b) Correct. In general, raising the classification threshold reduces false positives, thus raising precision.

✓ Consider a classification model that separates email into two categories: "spam" or "not spam." If you raise the classification threshold, what will happen to recall? *1/1

- a) Always stay constant
- b) Always increase
- c) Always decrease or stay the same ✓

Feedback

c) Correct. Raising our classification threshold will cause the number of true positives to decrease or stay the same and will cause the number of false negatives to increase or stay the same. Thus, recall will either stay constant or decrease.

✓ Consider two models (A and B) that each evaluate the same dataset. Which one of the following statements is true? *1/1

- (●) a) If model A has higher precision and higher recall rates than model B, then model A is probably better. ✓
- ( ) b) If model A has higher recall rates than model B, then model A is better.
- ( ) c) If Model A has higher precision rates than model B, then model A is better.

Feedback

a) Correct. In general, a model that outperforms another model on both precision and recall is likely the better model. Obviously, we'll need to make sure that comparison is being done at a precision / recall point that is useful in practice for this to be meaningful. For example, suppose our spam detection model needs to have at least 90% precision to be useful and avoid unnecessary false alarms. In this case, comparing one model at {20% precision, 99% recall} to another at {15% precision, 98% recall} is not particularly instructive, as neither model meets the 90% precision requirement. But with that caveat in mind, this is a good way to think about comparing models when using precision and recall.

✓ A sarcasm-detection model was trained on 80,000 text messages: 40,000 messages sent by adults (18 years and older) and 40,000 messages sent by minors (less than 18 years old). The model was then evaluated on a test set of 20,000 messages: 10,000 from adults and 10,000 from minors. The following confusion matrices show the results for each group (a positive prediction signifies a classification of "sarcastic"; a negative prediction signifies a classification of "not sarcastic"):
Which of the following statements about the model's test-set performance are true? *1/1

Adults

True Positives (TPs): 512	False Positives (FPs): 51
False Negatives (FNs): 36	True Negatives (TNs): 9401

$$\text{Precision} = \frac{TP}{TP + FP} = 0.909$$

$$\text{Recall} = \frac{TP}{TP + FN} = 0.934$$

Minors

True Positives (TPs): 2147	False Positives (FPs): 96
False Negatives (FNs): 2177	True Negatives (TNs): 5580

$$\text{Precision} = \frac{TP}{TP + FP} = 0.957$$

$$\text{Recall} = \frac{TP}{TP + FN} = 0.497$$

- [x] a) The model fails to classify approximately 50% of minors' sarcastic messages as "sarcastic." ✓
- [x] b) The 10,000 messages sent by adults are a class-imbalanced dataset. ✓
- [ ] c) Approximately 50% of messages sent by minors are classified as "sarcastic" incorrectly.
- [x] d) Overall, the model performs better on examples from adults than on examples from minors. ✓

Feedback

a) Correct. The recall rate of 0.497 for minors indicates that the model predicts "not sarcastic" for approximately 50% of minors' sarcastic texts.

b) Correct. If we compare the number of messages from adults that are actually sarcastic (TP+FN = 548) with the number of messages that are actually not sarcastic (TN + FP = 9452), we see that "not sarcastic" labels outnumber "sarcastic" labels by a ratio of approximately 17:1.

d) Correct. The model achieves both precision and recall rates over 90% when detecting sarcasm in text messages from adults. While the model achieves a slightly higher precision rate for minors than adults, the recall rate is substantially lower for minors, resulting in less reliable predictions for this group.

✓ A confusion matrix is a table used to evaluate the performance of a machine learning model by comparing predicted and actual values. *1/1

- (●) True ✓
- ( ) False

✓ The F1-score is a metric that combines precision and recall to evaluate the performance of a machine learning model. *1/1

- (●) True ✓
- ( ) False

✓ Precision and recall are both metrics used to evaluate the performance of a machine learning model. *1/1

- (●) True ✓
- ( ) False

✓ Which of the following statements accurately reflects why precision is more important than recall in email spam filtering? *1/1

- (●) False positives can cause significant problems, such as missed business opportunities or lost communication with important contacts. ✓
- ( ) False negatives can cause severe consequences in email spam filtering.
- ( ) None of the above.

✓ In a medical screening test for a severe disease, which of the following is more important to minimize *1/1

- ○ False positives
- ◉ False negatives ✓
- ○ Both are equally important