School of Computing and Information Systems
The University of Melbourne
COMP30027 MACHINE LEARNING (Semester 1, 2019)

Tutorial exercises: Week 10

1. What is the difference between "model bias" and "model variance"?
   (a) Why is a high bias, low variance classifier undesirable?
   (b) Why is a low bias, high variance classifier (usually) undesirable?
2. Describe how validation/development set, and cross-validation can help reduce overfitting?
3. Why `ensembling` reduces model variance?

1. What is the difference between "model bias" and "model variance"?

   (a) Why is a high bias, low variance classifier undesirable?

   (b) Why is a low bias, high variance classifier (usually) undesirable?

model bias: the propensity (trend) of a model to make same errors.

if no error → unbiased

if different kind of errors → unbiased.

in Regression: we can measure the difference between true value and target value.

in classification: we can only say same / different, so in classification to measure bias is to look the distribution of predict' classed doesn't match the actual class.

model variance: the propensity of a model to produce different classification a measure of inconsistency.

(a) Consistently wrong. the distribution of predicted class is consistently different with true class. this means that must make mistake.

(b) low bias → means making a bunch of correct prediction, but high variance → means not all of the predictions can possibly be correct, the correct prediction will change a lot when we change training data.

low bias → the distribution of predicted class is same as the distribution of actual class.

However, high variance → the instances assigned to one label may change next time.

2. Describe how validation/development set, and cross-validation can help reduce overfitting?

(see solution)

train on train → overfitting

So we use val/dev set to measure performance

if data set not so big, use cross validation

3. Why `ensembling` reduces model variance?

averaging reduce variance.

$$\mathrm{Var}\left(\frac{1}{N}\sum_i z_i\right) = \frac{1}{N}\mathrm{Var}(z_i)$$

the idea is if several model are average, the variance ↓, but no effect on bias.

There is only one training set. Ensembling creates multiple training set from one set (bagging, random forests) or by training multiple algorithm (Stacking). The prediction are combined to final result.