COMP47490 Tutorial

Evaluation in Machine Learning

1. The *confusion matrix* below shows the evaluation results for a binary classifier when applied to a test set of 768 examples, which are annotated with the class labels: (Pass, Fail).

Predicted Class

Pass	Fail	
407	93	Pass
108	160	Fail

Real Class

From this table calculate:

   a) The *precision* score for both of the classes.

   b) The *recall* score for both of the classes.

   c) The *F1-measure* score for both of the classes.

   d) The *overall classification accuracy* for the full test set.

2. The table below shows the true class labels for a test set of 12 emails, which are labelled as "spam" or "non-spam". The table also reports the predictions made by three different binary classifiers for those emails.

Example	True Class Label	KNN Prediction	J48 Prediction	SVM Prediction
1	spam	spam	spam	spam
2	non-spam	non-spam	spam	non-spam
3	spam	non-spam	non-spam	spam
4	non-spam	non-spam	non-spam	non-spam
5	spam	spam	spam	spam
6	non-spam	non-spam	non-spam	non-spam
7	non-spam	spam	spam	non-spam
8	non-spam	non-spam	spam	spam
9	spam	spam	non-spam	spam
10	spam	spam	non-spam	non-spam
11	spam	non-spam	non-spam	spam
12	spam	spam	spam	spam

a) Calculate the *overall accuracy* for each of the classifiers on this data. Based on your calculations, which classifier the most accurate?

b) Calculate the *precision* of each classifier relative to the "spam" class. Based on your calculations, which classifier achieves the highest precision for this class?

3. The table below shows the number of correct and incorrect predictions made by an image classifier during a 10-fold cross validation experiment, where the goal was to classify 500 images into one of three categories: {cats, dogs, people}.

Fold	Class: Cats		Class: Dogs		Class: People	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
1	82	68	82	68	164	36
2	81	69	102	48	176	24
3	99	51	97	53	160	40
4	81	69	102	48	148	52
5	94	56	99	51	148	52
6	97	53	91	59	162	38
7	81	69	94	56	148	52
8	76	74	79	71	181	19
9	76	74	97	53	160	40
10	96	54	79	71	179	21

a) What is the *overall accuracy* of the classifier based on the cross-validation results?

b) What conclusion might be draw about the different classes in the data, based on the results above?

c) Would *leave-one-out cross validation* be an appropriate evaluation strategy on this dataset?