Machine Learning

Seminar 4

1. A confusion matrix is an N dimensional square matrix, where N represents total number of target classes or categories. There are four important terms in a confusion matrix: TP, TN, FP, FN. Explain these terms.

	Predicted: NO	Predicted: YES
Actual: NO	True Negative (TN)	False Positive (FP)
Actual: YES	False Negative (FN)	True Positive (TP)

2. For the confusion matrix below, calculate the sensitivity, specificity, positive predictive value, negative predictive value and accuracy.

Confusion Matrix		Target	
		Positive	Negative
Model	Positive	70	20
	Negative	30	80

3. Below is a confusion matrix for a multi-class dataset. Convert it into a one-vs-all type matrix (binary-class confusion matrix).

		Expected			
		1	2	3	4
Predicted	1	52	3	7	2
	2	2	28	2	0
	3	5	2	25	12
	4	1	1	9	40

4. For the binary-class confusion matrix you derived in Q3, calculate the precision, recall and F1-score.

5. Hold-out cross-validation is the simplest and most common technique. You might not know that it is a hold-out method, but you certainly use it every day.

The algorithm of hold-out technique:

- Divide the dataset into two parts: the training set and the test set. Usually, 80% of the dataset goes to the training set and 20% to the test set but you may choose any splitting that suits you better
- Train the model on the training set
- Validate on the test set
- Save the result of the validation

What is the limitation of this result? Can you think of other better validation methods?