

**11482 Pattern Recognition and Machine Learning**  
**11512 Pattern Recognition and Machine Learning PG**

**Tutorial 4 – Week 5 Discussion Questions (Lab is attached separately)**

1. If you are working with training data that is full of errors and outliers, will this make it hard for the system to detect patterns? If so, what would you need to do to address this problem, so the system works properly?
2. Briefly describe 'regularisation' as used in regression-based classifiers? What purpose does it serve?
3. Show how logistic regression can be derived from linear regression. Take the normal regression equation, apply the logit, and you'll get out the logistic regression provided the criterion is binary.

So from

$$odds = \frac{P}{1-P}, \quad P = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

Where p is the probability of an event occurring, 1-p is the complementary that the event has not occurred.

Take a normal regression equation ( $y = b_0 + b_1 x$ ), apply the logit (logarithm of the odds of  $p/(1-p)$ ), and you'll get out the logistic regression. It is still linear in input data and 'normal regression' derived from linear regression.

**Show:  $\ln(p/(1-p)) = y$ ,**

where y can be any linear equation such as  $y = b_0 + b_1 x$

4. Consider a hypothetical classification problem for which a set of models' performance needs to be evaluated using the following outcomes.

The following table summarised the mean accuracy (mean standard deviation) for each model:

```

LR: 0.780037 (0.061129)
LDA: 0.773559 (0.058283)
KNN: 0.713432 (0.067707)
CART: 0.684109 (0.045221)
NB: 0.750820 (0.050575)
SVM: 0.750925 (0.063180)

```

a). Which model performs the best from the list? Why?

b). Results of predictions from a selected model from above gives the following predictions to the ground truth using the following code.

```

print(accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

```

```
0.7857142857142857
```

```
[[89  8]
 [25 32]]
```

	precision	recall	f1-score	support
0	0.78	0.92	0.84	97
1	0.80	0.56	0.66	57
avg / total	0.79	0.79	0.78	154

**Interpret the above classification report.**

5. Explain the problem of 'imbalanced classification'?  
Propose some strategies that you may use to resolve the imbalanced classification problem using an example discussed in lectures.