Tutorial 9 Solutions

1.

A) $TSS = \sum_o \quad \sum_{i=1}^{n}(y_i - \overline{y})^2$ where y_i is the desired output of node *i* and $\overline{y}$ the actual output.
The squared error for a particular training instance is the sum of squared errors taken over all output units and patterns.

B) $MSE = \frac{TSS}{n}$, The MSE (Mean Squared Error)
The average total squared error (Divided by the number of patterns).

C) Classification Error is the number of incorrectly classified instances divided by the total number of instances.

D) $MAE = \frac{\sum\sum_{i=1}^{n}|y_i - \bar{y}|}{n}$
Just average the magnitude of the individual errors without taking account of their sign.

2. To evaluate a classifier where the class is categorical (nominal) we use classification error. This usually given as a percentage but can be a fraction. In a neural network classifier TSS and MSE are NOT directly relevant but are needed as intermediate steps to the final result.

3. Duplicate question.

4. To assess error of a numeric predictor, that is the class is a number, we can use MAE or MSE. Classification error is NOT relevant (and it's not clear how you would calculate it).

5. We can use either of the above errors for train, test or validation set.

6. When MSE is large, the corresponding classification error will be large.

7. When MSE is small, the corresponding classification error will be small.

a. Since the data set is numeric no need for data encoding, but it would be a good idea to normalize/rescale to the range [0, 1].

b. tail -n +146 bodyfat.arff |cut -d, -f15 | sort -n | head -1
The minimum of this data is 4.1

c. tail -n +146 bodyfat.arff | cut -d, -f15|sort -n | tail -n -1
The maximum is 28.7

d. Min of this data is 4.1 and Max is 28.7.

e. We could either scale using min and max using the following formula or using theoretical boundaries.

$$y = \frac{X - \min(x)}{\max(y) - \min(y)}$$

f. $y = \frac{12.3 - 0}{100 - 0}$ = 0.123

g. $0.11 = \frac{X - 0}{100 - 0}$ → x= 11

h. 0 and 100

i. No clear choice

Using actual min/max will spread data evenly in the 0-1 range but can't predict a value outside the actual min/max. Using theoretical min/max will bunch the data at the lower end of 0-1 but it will be possible to predict outside the range.

j. Test accuracy will be all wrong if the scaling is different to train scaling.