Week 4 Answers

Q1

1. Book work
2. Entropy = 1
3. Odour – calculate the entropies of Mushroom|feature, and select lowest entropy (think about why this corresponds to maximum information gain)

Q2

1. Book work
2. Branches that split on feature X7 can be removed – you should make sure you understand why

Q3

TN =150, FP=30, TP = 60, FN = 10 – definitions of sensitivity, specificity, accuracy and F1 score are in the course notes.

Q4

The answer to this question is fairly involved. The steps are:

1. Realise which data points are important (the input features are irrelevant)
2. Based on the threshold, calculate the predicted class from the model output probability. When the threshold is zero, then ALL examples are classified to the positive class.
3. Calculate the true positive and false positive rates. For threshold = 0, TPR = 1, FPR = 1
4. Repeat the steps above for all thresholds, and plot TPR and FPR + join the dots.