Review Questions 2
ID2223
Group: Nicolas Oulianitski Essipova & Peter Lakatos

1. (a) True
   (b) False
   (c) True
   (d) False

2. The feature subset strategy determines, which attributes to use for the predictive models so that it yields the best accuracy. It can detect and remove irrelevant and redundant attributes, thus minimizing the size of the dataset and improving the complexity.

3. Entropy becomes zero when all class partitions are pure because that entails each partition only contains one class, hence pure. Calculation of entropy is as follows:

$$H_i = -\sum_{\substack{k=1 \\ p_{i,k} \neq 0}}^{n} p_{i,k} \log_2 (p_{i,k})$$

   So if each node only contains one class (i.e: $p_{i,k} = 1$ if input arrives at said node), then we find that entropy is zero.

4. With C total class and p(i) being the probability that we pick a datapoint with class i, then the Gini Impurity is:

$$G = \sum_{i=1}^{C} p(i) * (1 - p(i))$$

   When classes are pure, p(i) will be 1 and 0s. If it is 1 then 1-1=0 and if p(i) is 0 we also multiply by 0 and the result is 0.

5. Assuming we have a neural network with one hidden layer with linear functions as activation function for the hidden and output function, then the entire neural network ultimately collapses into a single layer – or more accurately; just becomes a linear algebraic transformation of the input's domain space into another. As per linear algebra: functions remain linear if all that is being done to the input is linear transformations (in this case; scaling and transforming from one coordinate space into another).

6. Deep feedforward neural networks require non-linear and differentiable activation functions. But the binary step function is not differentiable and the constant part yields a derivative of 0, while the step produces ∞.

7. Check email sent by Nicolas Essipova (nioe@kth.se) for photo on calculations