Exercises for Chapter 8, Part 2

8.1 Consider the data sets for two classes $X_1 = \{(0,0)\}$ and $X_2 = \{(1,0),(0,1)\}$.

b) Which discriminant line will be found by a support vector machine with linear kernels and infinite penalty for margin violations?

c) Which classification rule will be found by a nearest neighbor classifier?

d) Which classification rule will be found by a three nearest neighbor classifier?

e) Draw a receiver operating characteristic for these four classifiers, where class 2 is considered positive.

8.3 We build a support vector machine classifier for the XOR data set defined by $X_- = \{(-1,-1),(1,1)\}$ and $X_+ = \{(-1,1),(1,-1)\}$.

a) Are the classes of the XOR data set linearly separable?

b) We use the quadratic kernel $k(x,y) = (xy^T)^2$ and explicitly map the data to $\mathbb{R}^3$, where we use the squares of the original two-dimensional data as the first two dimensions. Determine the formula to compute the third dimension.

c) In the three-dimensional space from (b), determine the normal vector, offset, and margin of the separation plane that the support vector machine will find.

8.4 A car manufacturer has produced 50.000 cars with 90 kW gasoline engine, 100.000 cars with 120 kW gasoline engine, and 50.000 cars with 90 kW diesel engine. For all calculations use the following assumptions: $-\frac{1}{4}\log_2\frac{1}{4} = \frac{1}{2}$, $-\frac{1}{3}\log_2\frac{1}{3} = \frac{19}{36}$, $-\frac{1}{2}\log_2\frac{1}{2} = \frac{1}{2}$, $-\frac{2}{3}\log_2\frac{2}{3} = \frac{14}{36}$.

a) What is the entropy of these data?

b) For a given car, how much information do you gain if you know the kW value?

c) For a given car, how much information do you gain if you know whether it has a gasoline or diesel engine?

d) Sketch the ID3 decision tree.