EXAM TEST 13.12.17

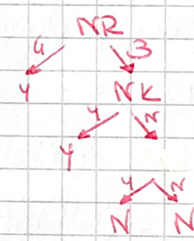
Question 1

1). This is a binary classification task. The target function is $f: X \to \{Y, N\}$ with $X = \{ F \times NR \times NK \}$ and the dataset is $D = \{(x_i, y_i)_{i=1}^{5}\}$.

2). We want to maximize the information gain: $IG(S, a) = ent(S) - \sum_{v \in V} \frac{|S_v|}{|S|} ent(S_v)$

with $ent(S) = -\sum_{i=1}^{N} p_i \log_2 p_i$ entropy. The entropy measures the impurity of our information. We define V like the set of values of A.

3). $IG(S, F) = -\frac{3}{5} ent(F_N) - \frac{2}{5} ent(F_Y) + ent(S) = 0.971 - 0.4 - 0.549 = 0.022$



$ent(F_N) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.916$

$ent(F_Y) = 1$

$ent(S) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$

$IG(S, NR) = ent(S) - \frac{3}{5} ent(NR_3) - \frac{2}{5} ent(NR_4) = 0.971 - 0.549 = 0.422$

HIGHEST IG!

$ent(NR_4) = 0 \qquad ent(NR_3) = 0.916$

$IG(S, NK) = -\frac{4}{5} ent(NK_N) - \frac{1}{5} ent(NK_Y) = 0.971 - 0.8 = 0.171$

$ent(NK_N) = 1 \qquad ent(NK_Y) = 0$

$IG(S, F) = -\frac{2}{3} ent(F_N) - \frac{1}{3} ent(F_Y) + ent(S) = 0.971 - 0.666 = 0.305$

$ent(F_N) = 1 \qquad ent(F_Y) = 0$

$IG(S, NK) = ent(S) - \frac{2}{3} ent(NK_N) - \frac{1}{3} ent(NK_Y) = 0.971$

$ent(NK_N) = 0 \qquad ent(NK_Y) = 0$

Question 2

1). Given a classification problem with a dataset D and an hypotheses space H we are interested in $p(H|D)$. We can compute (knowing that $h \in H$) $p(h|D)$.

$p(h|D) = \frac{p(D|h)p(h)}{p(D)}$. Now we define the Maximum A Posteriori (MAP) hypothesis.

$h_{MAP} = \arg\max_{h \in H} \frac{p(D|h)\,p(h)}{p(D)} = \arg\max_{h \in H} p(D|h)p(h)$. If we assume that $p(h_i) = p(h_j)$

we can simplify and we obtain the Maximum Likelihood (ML) hypothesis:

$h_{ML} = \arg\max_{h \in H} p(D|h)$

2). Is false. Now we do an example:

$h_1(x) = +$ $\quad h_2(x) = -$ $\quad h_3(x) = -$

$p(D|h_1) = 0.4$ $\quad p(D|h_3) = p(D|h_2) = 0.3$

$h_{ML} = \arg\max_{h \in H} p(D|h) = \arg\max \{0.4, 0.3, 0.3\} = +$

Now we introduce the concept of Bayes Optimal Classifier (BOC). BOC is an optimal classifier, in fact returns always the optimal solution.

$V_{OC} = \arg\max_{v \in V} \sum_{h_i \in H} p(v|x, h_i)p(h_i|D)$

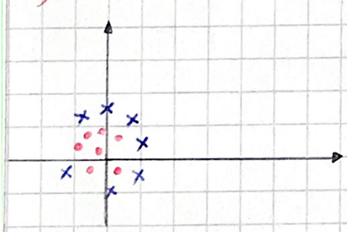
$p(+|x,h_1) = 1$ $\quad p(+|x,h_2) = 0$ $\quad p(+|x,h_3) = 0$

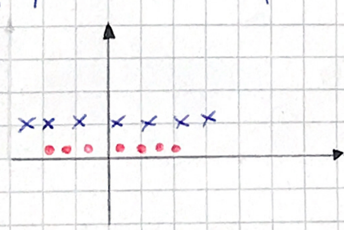
$p(-|x,h_1) = 0$ $\quad p(-|x,h_2) = 1$ $\quad p(-|x,h_3) = 1$

$V_{OC} = \arg\max_{v \in V} \{(1 \cdot 0.4 + 0 + 0), (0 + 1 \cdot 0.3 + 1 \cdot 0.3)\} = -$

Question 3

1). In this case i will use a polynomial kernel function.





2). We introduce some slack variables $\xi_n \geq 0$ $\forall n = 1 \dots N$

The modified error function becomes: $E(w) = \arg\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N} \xi_n$

And we modify also: $t_n y(x_n) \geq 1 - \xi_n$ $\forall n = 1 \dots N$

Question 4

1). This is a regression task, a suitable activation function can be the identity activation function $y = W^T h + b$

2). Using this activation function outputs units don't saturate so the learning step is not affected.

Question 5

1). $W_{OUT} = \frac{W_{IN} - W_K + 2p}{S} + 1 \longrightarrow W_{OUT} = \frac{3 - W_K + 2p}{S} + 1$

We choose stride $= 1$, padding $= 1$ and a 3×3 kernel.

2). tanh, ReLU, Sigmoid

Question 6