★ Spring 2012 Midterm

Problem 1

1] False. Naive Bayes is a linear classifier only in a particular case when we assume class independent variance.

2] False.
Classifier A overfits the training data & that is why it has lower test accuracy. ~~classifier~~ performance is measured in terms of overall accuracy i.e test set accuracy.

3] False.
Perceptron rule updates error for each wrong classification. Gradient descent is based on total reduction is squared error.

4] False. LR & Naive Bayes can represent even boolen feature concepts.

5] False.

D Tree can have 2^n nodes, where n is number of features.
Perceptron uses n weights which is smaller.

Problem 2]

1] $\frac{\partial Loss}{\partial w_1} = -\frac{5}{m} \sum_{i=1}^{m} (y_i - w_i x_i - w_0)^4 x_i$

2] (Online is not in syllabus i guess so I am implementing stochastic instead) which is similar i guess

=> Loop Until convergence
    for i in training dataset
        $w_1 = w_1 - \frac{5}{m}\alpha(y_i - w_i x_i - w_0)^4 x_i$
    end for
end loop

* α is learning rate

3] • stocastic gradient descent converges faster as we can comput Loss at every step.
• Batch gradient is more stable as it calculates loss over entire dataset

Problem 3

1] $F_1 = a$, $F_2 = c$, $F_3 = b$

$P(+) = \frac{2}{5}$, $P(-) = \frac{3}{5}$

$P(F_1 = a \mid +) = \frac{1}{2}$, $P(F_1 = a \mid -) = \frac{1}{3}$

$P(F_2 = c \mid +) = \frac{1}{2}$, $P(F_2 = c \mid -) = \frac{2}{3}$

$P(F_3 = b \mid +) = 0$, $P(F_3 = b \mid -) = \frac{1}{3}$

$P(+ \mid F) = 0$

$P(- \mid F) = \frac{3}{5} \times \frac{1}{\cancel{3}} \times \frac{2}{3} \times \frac{1}{3} = \frac{2}{45}$

2] Not taught

Problem 4

1] weight original $= \frac{1}{7}$

0 loss

Choose A, loss $= \frac{1}{7}$, $\alpha_m = \log(6)$

misclassified weight $= \frac{6}{7}$

Now,

take C as boundary

$$loss = \frac{\frac{2}{7}}{\frac{13}{7}} = \frac{2}{13}$$

take B as boundary

$$loss = \frac{\frac{1}{7}}{\frac{13}{7}} = \frac{1}{13}$$

∴ It will choose B

2] c] use smaller value of $\eta = 0.1$ because it seems that $\eta = 0.3$ increases $J(\theta)$ faster & instead of reaching optimum it might be surpassing it

3] • True. MLE is same but MAP is NOT

• False. MAP of thumback is larger

$$MAP\ coin = \frac{100 + 60 - 1}{300 - 2} = \frac{159}{298}$$

$$MAP\ thumback = \frac{1 + 60 - 1}{102 - 2} = \frac{60}{100} = \frac{3}{5}$$

* Problem 5 & 6 out of syllabus