I. Learning Methods

1.Linear Regression should be the best.

This is a regression problem to predict a numeric value instead of classification. Logistic regression, Decision tree, SVM and even perceptron are classification methods. We should pick Linear Regression method to predict Y = average rainfall in NYC based on X1, X2 = currents and tides in the Atlantic Ocean

2. C

3. B, C , (A :Since it's a typo, "The basic K-means algorithm requires setting up the parameter K (number of clusters) a prior**" is true while "**In K-means, we assume that each cluster fits a Gaussian distribution (normal distribution)" **is false)**

II. Naive Bayes

We would like to predict boy/girl gender based on the kids name and physical features, with an emphasis on unisex names.

1. Build a naive Bayes classifier using simple probabilities (e.g. no m-estimate or smoothing).

$$y_{new} = argmax_{y \in Y} p(y) \prod_j p(a_j|y)$$

Where $y = Sex: boy/girl$

$$a_1 = Name: Tyler/Salma/John/Leila/Alexandra/Ali,$$

$$a_2 = Tall: yes/no, a_3 = Eye: blue/brown, a_4 = Hair: short/long$$

$P(a_i\|y)$	$y = boy$	$y = girl$
$a_1 = Name: Tyler$	$1/3$	$2/5$
$a_1 = Name: Salma$	0	$1/5$
$a_1 = Name: John$	$1/3$	0
$a_1 = Name: Leila$	0	$1/5$
$a_1 = Name: Alexandra$	0	$1/5$
$a_1 = Name: Ali$	$1/3$	0
$a_2 = Tall: yes$	$2/3$	$2/5$
$a_2 = Tall: no$	$1/3$	$3/5$
$a_3 = Eye: blue$	$2/3$	$3/5$
$a_3 = Eye: brown$	$1/3$	$2/5$
$a_4 = Hair: short$	$2/3$	$1/5$
$a_4 = Hair: long$	$1/3$	$4/5$

$$P(boy) = \frac{3}{8}, P(girl) = \frac{5}{8}$$

2. What is the prediction for a "Tall kid named Tyler with long hair and brown eyes"? Justify briefly.

$$y_{new} = argmax_{x \in Y} p(y) \prod_j p(a_j|y)$$

$$= argmax_{y \in Y} p(y) * \text{p(Tyler|y)} * \text{p(Tall|y)} * \text{p(brown|y)} * \text{p(long | y)]}$$

$$boy = \frac{3}{8} * \frac{1}{3} * \frac{2}{3} * \frac{1}{3} * \frac{1}{3} = \frac{1}{108} \approx 0.01$$

$$girl = \frac{5}{8} * \frac{2}{5} * \frac{2}{5} * \frac{2}{5} * \frac{4}{5} = \frac{4}{125} \approx 0.032 > 0.01$$

y = girl.

So, the prediction is girl.

3. Based on our knowledge of frequencies in the larger population, we know the priors P (boy) = P(girl) ≈ 0.5. What are the priors based on the frequency in this training set? What can you say about this dataset? Explain.

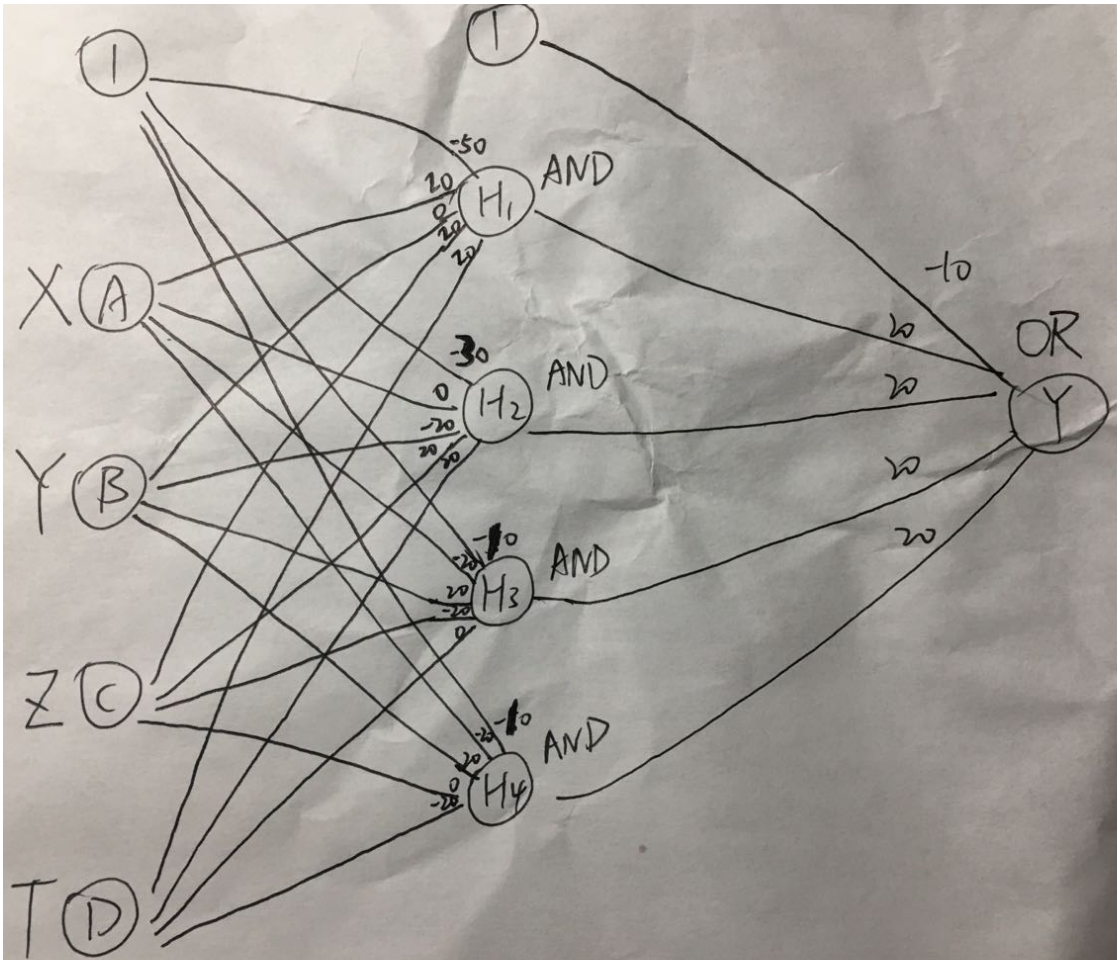
$$priors\ in\ this\ dataset: P(boy) = \frac{3}{8}, P(girl) = \frac{5}{8}$$

The dataset is too small to closely represent the prior and likelihood in large population. Especially the prior is biased to girl for all predictions causing more wrong prediction than using priori 0.5 based on knowledge of larger population.

III. Neural Networks

Construct a one-hidden layer neural network for the Boolean function below. Show all your work.

(X or not Y) XOR (not Z or not T)



$$A\ XOR\ B\ = [A\ AND\ (NOT\ B)]\ OR\ [(NOT\ A) AND\ B]$$

$$(X\ or\ not\ Y) XOR\ (not\ Z\ or\ not\ T) =$$

$$(A\ AND\ Z\ AND\ T) OR \big((NOT\ Y)\ AND\ Z\ AND\ T\big) OR \big((NOT\ X) AND\ Y\ AND (NOT\ Z)\big) OR ((NOT\ X) AND\ Y\ AND (NOT\ T))$$

$$g(X,Y,Z,T) = g(-10 + 20g(-50 + 20X + 20Z + 20T) + 20g(-30 - 20Y + 20Z + 20T) + 20g(-10 - 20X + 20Y - 20Z) + 20g(-10 - 20X + 20Y - 20T))$$

[BONUS +3]

Consider a simple neural network model with one linear output neuron and no hidden layer. We want to take the average of two such networks. In other words, given input and two networks with weights respectively, the output is:

$$y = \frac{1}{2}\mathbf{w_1}^T\mathbf{x} + \frac{1}{2}\mathbf{w_2}^T\mathbf{x}$$

Will this improve the performance of this model? Prove your claim.

Yes. This method can be viewed as model averaging which is a kind of very simple ensemble method. Two simple neural network with different weights can be viewed as weak learners and thus the average of their prediction can be better than a single one.