HW3: Linear Regression
CS383 - Machine Learning
Duy Le

Part 1.1: Theory

$$H(a,b) = -\frac{a}{a+b}log_2\frac{a}{a+b} - \frac{b}{a+b}log_2\frac{b}{a+b}$$

A.

$$Y\_positive = 3 + 4 + 4 + 1 = 12$$
$$Y\_negative = 0 + 1 + 3 + 5 = 9$$

$H(Y) = H(9,12) = 0.985$

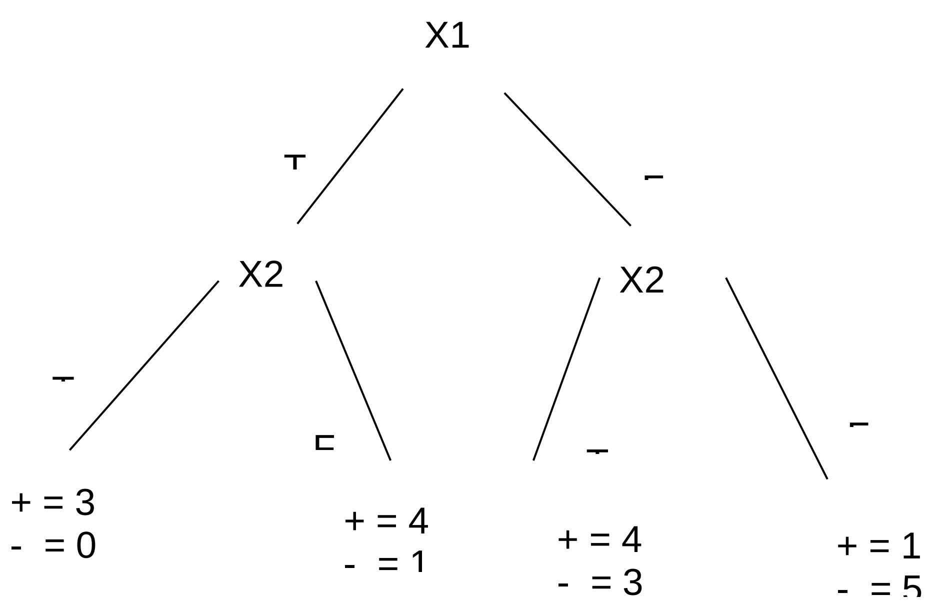
B.
Feature X1:

$$P_T = 7, N_T = 1$$
$$P_F = 5, N_F = 8$$
$$E(H(X1)) = \frac{8}{21}H(7,1) + \frac{13}{21}H(5,8) = 0.20707 + 0.59505 = 0.802$$
$$IG(X1) = H(Y) - E(H(X1)) = 0.985 - 0.802 = 0.183$$

Feature X2:

$$P_T = 7, N_T = 3$$
$$P_5 = 5, N_F = 6$$
$$E(H(X1)) = \frac{10}{21}H(7,3) + \frac{11}{21}H(5,6) = 0.420 + 0.5207 = 0.940$$
$$IG(X1) = H(Y) - E(H(X1)) = 0.985 - 0.940 = 0.045$$

C.



Part 1.2: Theory

A.

$$P(A = Yes) = 3/5$$
$$P(A = No) = 2/5$$

B.

X =
[[216. 5.68]
[69. 4.78]
[302. 2.31] => Standardize =>
[60. 3.16]
[393. 4.2]]

X =
[[0.05509059 1.24771393]
[-0.95719904 0.56878857]
[0.64731446 -1.29448434]
[-1.01917595 -0.65327706]
[1.27396994 0.1312589]]

mean = [208 4.026]
Stdev = [145.2 1.326]

Class A:
[[0.05509059 1.24771393]
[-0.95719904 0.56878857]
[-1.01917595 -0.65327706]
mean_A = [-0.640 0.387]
Stdev_A = [0.603 0.963]

Class non A:
[0.64731446 -1.29448434]
[1.27396994 0.1312589]
mean_nonA = [0.961 -0.582]
Stdev_nonA = [0.443 1.008]

C.

$$Gaussian(x_k, \mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(x_k-\mu_i)^2}{2\sigma_i^2}}$$

$$F = [242\ \ 4.56] => standardize => F = [0.234\ \ \ 0.403]$$

$$P(Y = A) = P(A = Yes)\, Gaussian(0.234, -0.640, 0.603)\, Gaussian(0.403, 0.387, 0.963)$$
$$= 0.6 * 0.231 * 0.414 = 0.0574$$
$$P(Y \,!= A) = P(A = No)\, Gaussian(0.234, 0.961, 0.443)\, Gaussian(0.403, -0.582, 1.008)$$
$$= 0.4 * 0.234 * 0.2455 = 0.0229$$

P(Y = A) > P(Y != A) => That student get an A

Part 1.3 Theory

A.

Since g(x) has to be in range [0,1] but tanh(x) is in range [-1,1] =>

$$g(x\theta) = \frac{tanh(x\theta) + 1}{2}$$

B.

$$tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

Proof: $\frac{\partial}{\partial\theta_j}(tanh(x\theta)) = x_j(1 - tanh(x\theta)^2).$

Solve:

Note $\frac{d}{d\theta_j}\ e^{x\theta}\ =\ x_j e^{x\theta}$ and $\frac{d}{d\theta_j}\ e^{x\theta}\ =\ x_j e^{x\theta}$

$$\frac{d}{d\theta_j} tanh(x\theta) = \frac{\frac{d}{d\theta_j}(e^{x\theta} - e^{-x\theta})(e^{x\theta} + e^{-x\theta}) - (e^{x\theta} - e^{-x\theta})\frac{d}{d\theta_j}(e^{x\theta} + e^{-x\theta})}{(e^{x\theta} + e^{-x\theta})^2}$$

$$\frac{d}{d\theta_j} tanh(x\theta) = \frac{(x_j e^{x\theta} + x_j e^{-x\theta})(e^{x\theta} + e^{-x\theta}) - (e^{x\theta} - e^{-x\theta})(x_j e^{x\theta} - x_j e^{-x\theta})}{(e^{x\theta} + e^{-x\theta})^2}$$

$$\frac{d}{d\theta_j} tanh(x\theta) = x_j \frac{(e^{x\theta} + e^{-x\theta})(e^{x\theta} + e^{-x\theta}) - (e^{x\theta} - e^{-x\theta})(e^{x\theta} - e^{-x\theta})}{(e^{x\theta} + e^{-x\theta})^2}$$

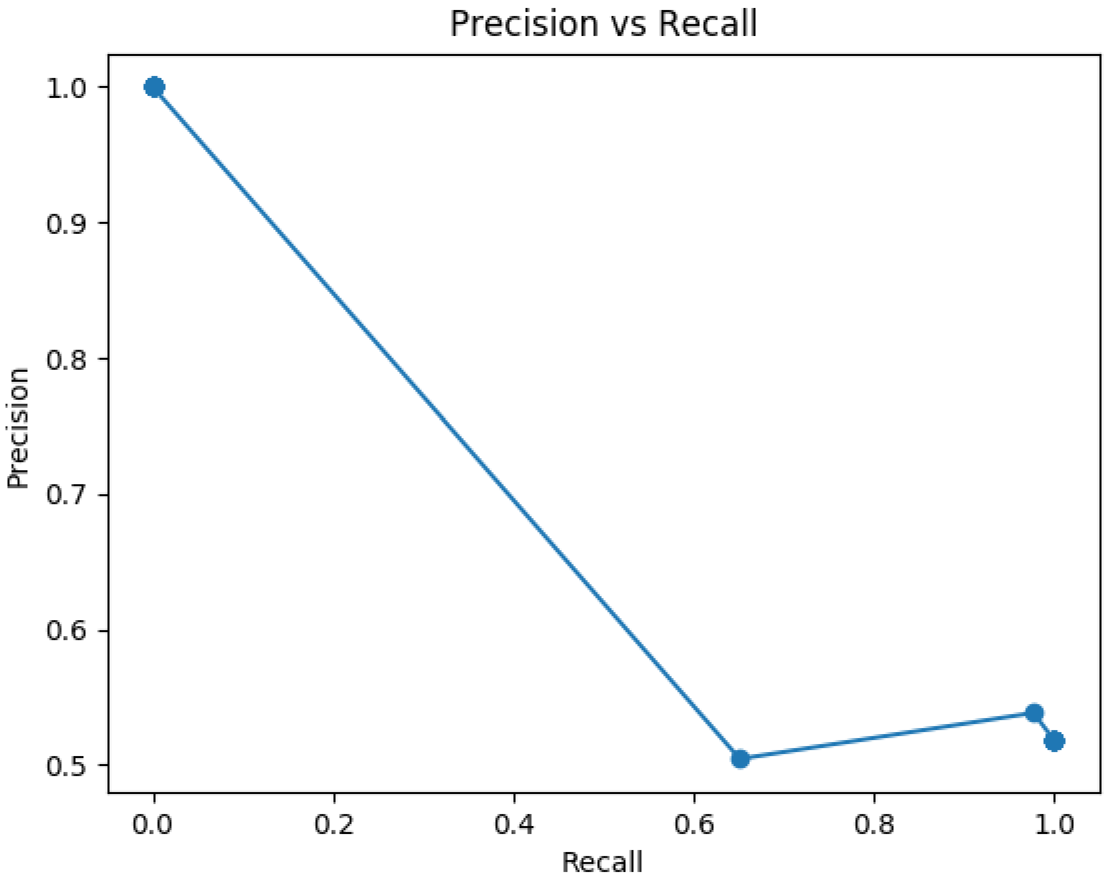
$$\frac{d}{d\theta_j} tanh(x\theta) = x_j(1 - \frac{(e^{x\theta} - e^{-x\theta})^2}{(e^{x\theta} + e^{-x\theta})^2}) = x_j(1 - tanh(x\theta)^2)$$

Part 2: Naive Bayes

Precision: 0.719660
Recall: 0.962662
F-Measure: 0.823611
Accuracy: 0.834312

Part 3: Recall and Precision:

Dataset:
https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex/version/2?#500_Person_Gender_Height_Weight_Index.csv
This dataset was modified to be similar to given format (no header and last column is label) in order to be used in our function



Precision vs Recall
1.0
0.9
0.8
0.7
0.6
0.5
Precision
0.0
0.2
0.4
0.6
0.8
1.0
Recall