Q1
Assumptions of Naive Bayes Classifier:
Attribute values are conditionally independent given the target value. The assumption is that given the target value of the instance, the probability of observing the conjunction $a_1, a_2.. .a_n$, is just the product of the probabilities. This assumption simplifies the computation of classifying the new instance to assign the most probable target value, v_{MAP}, given the attribute values ($a_1, a_2.. .a_n$).

(Assumption:2 Advantage:2)

X= {Outlook=Rainy, temperature=Cool, Humidity=High, Wind=Yes}

Prior probability:
P(play=Yes)=6/10
P(play=No)=4/10

Conditional probability:
P(Outlook=Rainy | play=Yes) = 1/6
P(temperature=Cool | play=Yes) = 3/6
P(Humidity=High | play=Yes) = 1/6
P(Wind=Yes | play=Yes) = 1/6

P(Outlook=Rainy | play=No) = 3/4
P(temperature=Cool | play= No) = 1/4
P(Humidity=High | play= No) = 3/4
P(Wind=Yes | play= No) = 2/4

(Prior and conditional probabilities: 10 (1 each))

P(play=Yes | X) = 1/6*3/6*1/6*1/6*6/10 = 0.001389

P(play = No|X) = 3/4*1/4*3/4*2/4*4/10 = 0.028125

MAP = max (0.001389, 0.028125)

Hence the decision is *Don't play*

(Posterior probabilities, MAP and prediction: 4.75)

Q2

Height (x)	63	64	66	69	69	71	71	72	73	75
Weight (y)	127	121	142	157	162	156	169	165	181	208

Mean(x) = 69.1666667
Variance(y) = 160.666667
Covariance = 109.866667
Variance = 17.7666667
The student could have used any formula.
coefficient (m) = 6.18386
coefficient (c) = - 267.05066

$\hat{y} = 6.18386\,x - 267.05066$

(Derivation of m, c and the fit $\hat{y}$: 10 marks)

y	127	142	162	156	169	208
yhat	122.532833	141.084428	159.636023	172.003752	172.003752	196.739212
y-yhat	4.46716698	0.91557223	2.36397749	-16.003752	-3.0037523	11.260788
square(y-yha	19.9555808	0.83827251	5.58838955	256.120089	9.02252815	126.805346

Each value (y-yhat) is the ith residual.

MSE = 69.7217011

(Residual: calculation of y-yhat for each point or square(y-yhat) or MSE or MAE: 6)

For x = 67, y = 147.268293

(prediction: 2.75)

Q3 For the following multivariate ML problem, the initial values for the regression coefficients β_0, $\beta_1, and\ \beta_2$ are 1, 1, and 1. Find the initial cost of the problem. Given the learning rate $\alpha = 0.10$, compute the next set of values for the regression coefficients $\beta_0,\ \beta_1, and\ \beta_2$. What is the cost of the function on the new values of the coefficients?

X1	X2	Y
0	1	4
1	2	7
2	2	8
3	1	7
2	1	6

Ans. Iteration 1, Starting $J = \frac{1}{2m}\sum(\beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 - Y)^2$

$$= \frac{1}{2.5}\left[(2-4)^2 + (4-7)^2 + (5-8)^2 + (5-7)^2 + (4-6)^2\right]$$

$$= \frac{1}{2.5}\left[4+9+9+4+4\right] = \frac{1}{2.5}\cdot 30 = 3$$

$$\beta_0 = \beta_0 - \frac{0.1}{5}\left[(-2)1 + (-3).1 + (-3).1 + (-2).1 + (-2).1\right]$$

$$= 1 - \frac{0.1}{5}\left[-2-3-3-2-2\right] = 1 + \frac{1.2}{5} = 1 + 0.24 = 1.24$$

$$\beta_1 = 1 - \frac{0.1}{5}\left[(-2).0 + (-3).1 + (-3)2 + (-2)3 + (-2)2\right]$$

$$1 + \frac{0.1}{5}\left[-3-6-6-4\right]$$

$\beta_1 = \ldots\ldots$

or Directly $\boxed{\text{Theta} - \frac{\alpha}{m}\left[X^T(X\text{theta} - Y)\right]}$

$$\beta = \begin{bmatrix}1\\1\\1\end{bmatrix} - \frac{0.1}{5}\begin{bmatrix}1&1&1&1&1\\0&1&2&3&2\\1&2&2&1&1\end{bmatrix}\left(\begin{bmatrix}1&0&1\\1&1&2\\1&2&2\\1&3&1\\1&2&1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} - \begin{bmatrix}4\\7\\8\\7\\6\end{bmatrix}\right)$$

$$= \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{0.1}{5}\begin{bmatrix}1&1&1&1&1\\0&1&2&3&2\\1&2&2&1&1\end{bmatrix}\begin{bmatrix}-2\\-3\\-3\\-2\\-2\end{bmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{0.1}{5}\begin{pmatrix}-12\\-19\\-18\end{pmatrix}$$

$$= \begin{pmatrix}1+\frac{1.2}{5}\\1+\frac{1.9}{5}\\1+\frac{1.8}{5}\end{pmatrix} = \begin{matrix}1+0.24\\1+0.38\\1+0.36\end{matrix} = \begin{matrix}1.24\\1.38\\1.36\end{matrix}$$

$$\text{New } J = \frac{1}{2m}\sum(X\theta - Y)^2 = \frac{1}{2.5}\left[1.96 + 3.0976 + 0.0784 + 0.4096\right]$$

$$= \frac{1}{10}(5.5456) = \boxed{0.55456}$$

Initial cost J calculation: 3
Derivation of each of the 3 coefficients: 3*4=12
New J calculation: 3.75

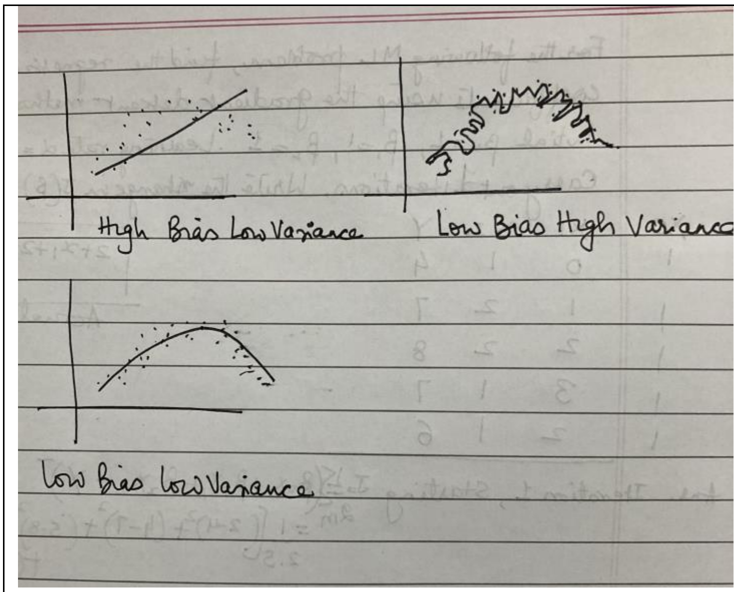
Q4
Regularization is the approach which involves fitting a model involving all predictors. The estimated coefficients are shrunk towards zero relative to the least squares estimates. It has the effect of reducing variance. It can also perform variable selection.

a) The regularization term is zero.
   Ans: The penalty term has no effect. Model is as complex as it was, overfitting
b) The regularization term is very large.
   Ans: The impact of penalty term grows and the coefficients approach zero. The model becomes linear, underfitting

(Regularization:3, a): 3 b): 3)

With the help of non-linear data, graphically show the following scenarios:
a) High Bias low variance
b) Low bias high variance
c) Low bias low variance

(a), b) and c) 9.75)



Q5

Which cost function is used for the logistic regression? Why is the mean squared error not suitable as a cost function in logistic regression? Certain health risk factors such as high blood pressure and cigarette smoking etc. lead to sudden death. Therefore a multiple logistic regression was fitted to these data as shown below.

Risk Factor	Regression Coefficient
Constant term	-15.3
Blood Pressure (mm Hg)	.099
Weight (Kg)	-.0060
Cholesterol (mg/100 mL)	.0056
Glucose (mg/100 mL)	.0066
Smoking (cigarettes/day)	.0069
Age (years)	.0686

Predict sudden death for a 50 year old man with diastolic blood pressure of 120 mmHg, a relative weight of 100 Kg of study mean, a cholesterol level of 250 mg/100mL, a glucose level of 100 mg/100mL who smokes 10 cigarettes per day. Also, predict the death if diastolic blood pressure is 180 mmHg.

Answer:

Cost function of logistic regression:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost(h_\theta x^i, y^i) \ where\ cost(h_\theta(x),\ y) = -y\log h_\theta(x) - [(1-y)\log(1-h_\theta(x))]$$

Cost function: 3.75

Since the hypothesis function for logistic regression is expressed as the non-linear sigmoid function (logistic function), if the mean square error is considered as the cost function, it is a non-convex function with many local minima. Hence the $J(\theta)$ ends up being non-convex making it unsuitable as a cost function for logistic regression.

Why not mean square error: 3

So the aim is to come up with a cost function which is convex given our hypothesis of the sigmoid function.

Numerical:

Plugging in the given values will give us the log odds of sudden death for such a man:

p = -15.3+.099*(120)-.006*(100)+.0056*(250)+.0066*(100)+.0069*(10)+ 0.0686*(50)

p = 1.5390000000000001

To get the actual probability of sudden death we note that

1/(1+np.exp(-p)) = 0.8233193076240006

(prediction of probability : 6)

For BP=180

p = -15.3+.099*(180)-.006*(100)+.0056*(250)+.0066*(100)+.0069*(10) + 0.0686*(50)

p = 7.478999999999999

The probability of death= 1/(1+np.exp(-p)) = 0.9994354968885402

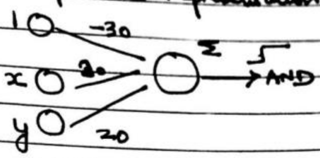
(prediction of probability : 6)

Q6

Boolean function AND

x	y	x and y
0	0	0
0	1	0
1	0	0
1	1	1

AND (one possible representation of weights)

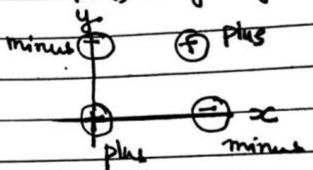


any other weights (-1, 0.8, 0.8) etc. are also correct

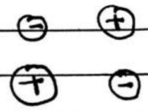
Consider a non-linearly separable function XNOR

x	y	XNOR
0	0	1
0	1	0
1	0	0
1	1	1

$f(x,y) = xy + x'y'$



A single perceptron cannot represent this non-linearly separable function, as can be seen in the figure.



We will have to combine 3 perceptrons, AND, NOR, OR

Scanned with CamScanner

AND truth table and possible architecture:10

Suitable illustration of non-linearity: 8.75