Note: This document is not a preview of what will appear on your exam, but a demonstration of how the different methods and formulas we've discussed in class can be used.

K-Nearest Neighbors

<u>Dataset with categorical label</u>

Temperature	Humidity	Weather
5	14	sunny
20	10	sunny
16	4	sunny
26	23	rainy
21	25	rainy
17	14	rainy
34	29	rainy

1. Predict the weather given temperature = 18 and humidity = 20 using Euclidean distance and k = 1

Temperature	Humidity	Distance from (18, 20)	Weather
5	14	$\sqrt{(18-5)^2 + (20-14)^2} = 14.32$	sunny
20	10	$\sqrt{(18-20)^2 + (20-10)^2} = 10.20$	sunny
16	4	$\sqrt{(18-16)^2 + (20-4)^2} = 16.12$	sunny
26	23	$\sqrt{(18-26)^2 + (20-23)^2} = 8.54$	rainy
21	25	$\sqrt{(18-21)^2 + (20-25)^2} = 5.83$	rainy
17	14	$\sqrt{(18-17)^2 + (20-14)^2} = 6.08$	rainy
34	29	$\sqrt{(18-34)^2 + (20-29)^2} = 18.36$	rainy

Answer:
5.83 → rainy

2. Predict the weather for the same data point, but with k = 3

Answer:
5.83 (rainy), 6.08 (rainy), 8.54 (rainy) → rainy (majority)

3. Normalize the features before predicting the weather for the same data point with k = 3

Temperature: min = 5, max = 34
Humidity: min = 4, max = 29

Normalize (18, 20)
$\frac{18-5}{34-5} = 0.45$ $\frac{20-4}{29-4} = 0.64$
(0.45, 0.64)

Temperature	Humidity	Distance from (0.45, 0.64)	Weather
$\frac{5-5}{34-5} = 0$	$\frac{14-4}{29-4} = 0.40$	$\sqrt{(0.45 - 0)^2 + (0.64 - 0.40)^2} = 0.51$	sunny
$\frac{20-5}{34-5} = 0.52$	$\frac{10-4}{29-4} = 0.24$	$\sqrt{(0.45 - 0.52)^2 + (0.64 - 0.24)^2} = 0.41$	sunny
$\frac{16-5}{34-5} = 0.38$	$\frac{4-4}{29-4} = 0$	$\sqrt{(0.45 - 0.38)^2 + (0.64 - 0)^2} = 0.64$	sunny
$\frac{26-5}{34-5} = 0.72$	$\frac{23-4}{29-4} = 0.76$	$\sqrt{(0.45 - 0.72)^2 + (0.64 - 0.76)^2} = 0.30$	rainy
$\frac{21-5}{34-5} = 0.55$	$\frac{25-4}{29-4} = 0.84$	$\sqrt{(0.45 - 0.55)^2 + (0.64 - 0.84)^2} = 0.22$	rainy
$\frac{17-5}{34-5} = 0.41$	$\frac{14-4}{29-4} = 0.40$	$\sqrt{(0.45 - 0.41)^2 + (0.64 - 0.40)^2} = 0.24$	rainy
$\frac{34-5}{34-5} = 1$	$\frac{29-4}{29-4} = 1$	$\sqrt{(0.45 - 1)^2 + (0.64 - 1)^2} = 0.66$	rainy

Answer:
0.22 (rainy), 0.24 (rainy), 0.30 (rainy) → rainy (majority)

Dataset with numerical target value

Temperature	Humidity	Rain Probability
5	14	0.10
20	10	0.20
16	4	0.15
26	23	0.82
21	25	0.91
17	14	0.73
34	29	0.95

4. Predict the rain probability given temperature = 18 and humidity = 20 using Euclidean distance and k = 3

Temperature	Humidity	Distance from (18, 20)	Rain Probability
5	14	$\sqrt{(18-5)^2+(20-14)^2} = 14.32$	0.10
20	10	$\sqrt{(18-20)^2+(20-10)^2} = 10.20$	0.20
16	4	$\sqrt{(18-16)^2+(20-4)^2} = 16.12$	0.15
26	23	$\sqrt{(18-26)^2+(20-23)^2} = 8.54$	0.82
21	25	$\sqrt{(18-21)^2+(20-25)^2} = 5.83$	0.91
17	14	$\sqrt{(18-17)^2+(20-14)^2} = 6.08$	0.73
34	29	$\sqrt{(18-34)^2+(20-29)^2} = 18.36$	0.95

Answer:
5.83 (0.91), 6.08 (0.73), 8.54 (0.82) → 0.82 (average of the corresponding rain probabilities)

Linear Regression

<u>Actual vs Predicted Values</u>

Actual House Prices (in million PHP)	Predicted House Prices (in million PHP)
4.08	4.02
4.12	4.19
5.35	5.21
6.68	6.75
7.24	7.39
8.35	8.23
9.80	9.94
9.92	9.78
10.36	10.45
10.64	10.52

1. Calculate the loss using Mean Squared Error (MSE)

$$\frac{1}{2n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = \frac{(4.08-4.02)^2+(4.12-4.19)^2+ ... + (10.64-10.52)^2}{2(10)} = 0.00658$$

2. Assuming that the coefficients of a linear regression model are given by $\theta = [0.19 \ \ 0.36]$ and the resulting gradient of the loss $\frac{\partial}{\partial\theta}l(\theta) = [0.02 \ \ 0.03]$, what are the resulting coefficients after one epoch of gradient descent with $\alpha = 0.01$? Answers should be in 4 decimal places.

$$\theta = \theta - \alpha\frac{\partial}{\partial\theta}l(\theta) = [0.19 \ \ 0.36] - (0.01)[0.02 \ \ 0.03] = [0.1898 \ \ 0.3597]$$

Logistic Regression

True Labels vs Predicted Probabilities

True Label	Predicted Probability
1	0.8
0	0.3
1	0.9
0	0.2
1	0.7

1. Assuming that the table above is the outcome of a binomial logistic regression model, compute its resulting loss value.

$$-\frac{1}{n}\sum_{i=1}^{n}(y_i log(z_i) + (1 - y_i)log(1 - y_i)) = \frac{-(log(0.8)+log(1-0.3)+log(0.9)+log(1-0.2)+log(0.7))}{5} = 0.2530$$

Note: The log() used here is natural logarithm, which is the ln() function

Multiple Classes

x (Feature)	Weights			y (True Values)		
	Class A	Class B	Class C	Class A	Class B	Class C
12	0.5	0.2	-0.1	1	0	0
31	0.2	0.4	0.1	0	1	0
24	0.4	-0.3	0.6	0	0	1
8	0.3	0	-0.4	1	0	0

Class	Bias
A	0.3
B	1.2
C	0.7

2. The tables above provide the inputs for a classification model, as well as the weights for the different classes and their respective biases. Predict the class of each instance using multinomial logistic regression.

x	Scores $z_i = x_i w_i + b$			Softmax		
	Class A	Class B	Class C	Class A	Class B	Class C
12	(12)(0.5) + 0.3 = 6.3	(12)(0.2) + 1.2 = 3.6	(12)(-0.1) + 0.7 = -0.5	0.9367	0.0623	0.0010
31	(31)(0.2) + 0.3 = 6.5	(31)(0.4) + 1.2 = 13.6	(31)(0.1) + 0.7 = 3.8	0.0008	0.9991	0.0001
24	(24)(0.4) + 0.3 = 9.9	(24)(-0.3) + 1.2 = -6.0	(24)(0.6) + 0.7 = 15.1	0.0055	0	0.9945
8	(8)(0.3) + 0.3 = 2.7	(8)(0) + 1.2 = 1.2	(8)(-0.4) + 0.7 = -2.5	0.8139	0.1816	0.0045

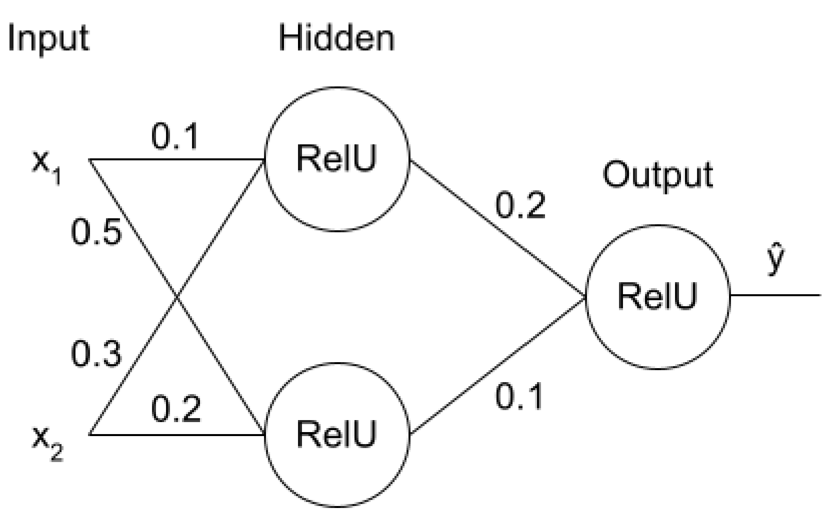
2. Compute the loss for this given example

$$-\frac{1}{n}\sum_{i=1}^{n} y_i^T log(p_i) = \frac{-(log(0.9367)+log(0.9991)+log(0.9945)+log(0.8139))}{4} = \frac{0.0654 + 0.0009 + 0.0055 + 0.2059}{4} = 0.0694$$

Notes:

- The log() used here is natural logarithm, which is the ln() function
- y_i^T is the vector of true values (one-hot encoded, ex: [1 0 0]) while p_i is the resulting predictions for a given instance (ex: [0.9367 0.0623 0.0010]). Since y_i^T contains one value that is equal to 1 and the rest are 0s, then $y_i^T log(p_i)$ is reduced to a single value

Neural Networks



Consider the neural network above with 2 layers, 2 input nodes and 1 output node. Both the hidden and output layer nodes use the RelU activation function. The weights are the values indicated in the connecting lines.

1. Given the test inputs (with their respective true values) below for 3 instances

i	x_1	x_2	y (True values)
1	24	17	3.2
2	15	33	4.1
3	8	11	1.8

What is the output of the neural network for each instance?

i = 1
$z_1 = (24)(0.1) + (17)(0.3) = 7.5$
$\text{Hidden}_1 = \text{RelU}(7.5) = \max(0, 7.5) = 7.5$

$z_2 = (24)(0.5) + (17)(0.2) = 15.4$
$\text{Hidden}_1 = \text{RelU}(15.4) = \max(0, 15.4) = 7.5$

$z_3 = (7.5)(0.2) + (15.4)(0.1) = 3.04$
$\hat{y} = \text{RelU}(3.04) = \max(0, 3.04) = 3.04$

i = 2
$z_1 = (15)(0.1) + (33)(0.3) = 11.4$
$\text{Hidden}_1 = \text{RelU}(11.4) = \max(0, 11.4) = 11.4$

$z_2 = (15)(0.5) + (33)(0.2) = 14.1$
$\text{Hidden}_1 = \text{RelU}(14.1) = \max(0, 14.1) = 14.1$

$z_3 = (11.4)(0.2) + (14.1)(0.1) = 3.69$
$\hat{y} = \text{RelU}(3.04) = \max(0, 3.04) = 3.69$

i = 3
$z_1 = (8)(0.1) + (11)(0.3) = 4.1$
$\text{Hidden}_1 = \text{RelU}(4.1) = \max(0, 4.1) = 4.1$

$z_2 = (8)(0.5) + (11)(0.2) = 6.2$
$\text{Hidden}_1 = \text{RelU}(6.2) = \max(0, 6.2) = 6.2$

$z_3 = (4.1)(0.2) + (6.2)(0.1) = 1.44$
$\hat{y} = \text{RelU}(1.44) = \max(0, 1.44) = 1.44$

2. What is the loss for this given example?

$$\frac{1}{2n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = \frac{(3.04-3.2)^2+(3.69-4.1)^2+(1.44-1.8)^2}{2(3)} = \frac{0.3233}{} = 0.0539$$

Naive Bayes

Student	Math grade	Hours studying	ML grade
0	Bad	>= 4hrs	passed
1	Good	>= 4hrs	passed
2	Bad	< 4hrs	failed
3	Bad	>= 4hrs	passed
4	Good	>= 4hrs	passed
5	Bad	>= 4hrs	failed
6	Good	< 4hrs	passed
7	Good	< 4hrs	failed
8	Good	>= 4hrs	passed
9	Bad	>= 4hrs	failed

Use Naive Bayes to answer the following questions about the dataset above

1. What is P(Pass | Good, >= 4 hrs)?

You only need this in general

$$P(Pass \mid Good,\ \ >=\ \ 4\ hrs) = \frac{P(Good \mid Pass)P(>= 4\ hrs \mid Pass)P(Pass)}{P(Good, >= 4\ hrs)} = \frac{(4/6)(5/6)(6/10)}{(5/10)(7/10)} = \frac{0.33}{0.35} = 0.94$$

In case you need resulting probability value

2. What is P(Fail | Good, >= 4 hrs)?

$$P(Fail \mid Good,\ \ >=\ \ 4\ hrs) = \frac{P(Good \mid Fail)P(>= 4\ hrs \mid Fail)P(Fail)}{P(Good, >= 4\ hrs)} = \frac{(1/4)(2/4)(4/10)}{(5/10)(7/10)} = \frac{0.05}{0.35} = 0.14$$

Let's say the previous dataset is updated as

Student	Math grade	Hours studying	ML grade
0	Bad	>= 4hrs	passed
1	Good	>= 4hrs	passed
2	Bad	< 4hrs	failed
3	Bad	>= 4hrs	passed
4	Good	>= 4hrs	passed
5	Bad	>= 4hrs	failed
6	Good	< 4hrs	passed
7	Bad	< 4hrs	failed
8	Good	>= 4hrs	passed
9	Bad	>= 4hrs	failed

Using MAP $\hat{\theta}_{MAP} = \frac{\alpha H + \beta H - 1}{\alpha H + \beta H - 1 + \alpha T + \beta T - 1}$ with $\beta H = 1$ and $\beta T = 1$, compute for the following

3. P(Good | Pass)

$$P(Good \mid Pass) = \frac{4 + 1 - 1}{4 + 1 - 1 + 2 + 1 - 1} = \frac{4}{6} = 0.67$$

Notes:

- αH = number of "successful" outcomes related to the prior parameter
  For $P(Good \mid Pass)$, $\alpha H = 4$ (good at math who passed ML)
- αT = number of "failure" outcomes related to the prior parameter
  For $P(Good \mid Pass)$, $\alpha T = 2$ (bad at math who passed ML)

4. P(Bad | Pass)

$$P(Bad \mid Pass) = \frac{2 + 1 - 1}{2 + 1 - 1 + 4 + 1 - 1} = \frac{2}{6} = 0.33$$

Note: In this case, the "successful" outcome is "bad at math who passed ML", thus $\alpha H = 2$ and the "failure" outcome is "good at math who passed ML", thus $\alpha T = 4$

Student	Math grade	Hours studying	ML grade
0	1.0	6	passed
1	4.0	5	passed
2	1.5	2	failed
3	2.0	6	passed
4	3.5	7	passed
5	1.5	4	failed
6	3.0	3	passed
7	4.0	2	failed
8	3.5	6	passed
9	1.0	4	failed

Use Naive Bayes to answer the following questions about the dataset above with continuous features

5. P(Pass | 3.5, 5)

Probability density function for normal distribution

$$pdf(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Math Grades of Passed = [1.0, 4.0, 2.0, 3.5, 3.0, 3.5]

$\mu\ (mean) = 2.8,\ \sigma\ (std\ dev) = 1.125$

$x = 3.5$

$$pdf(3.5, 2.8, 1.125) = \frac{1}{1.125\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{3.5-2.8}{1.125}\right)^2} = 0.292$$

Math Grades (All) = [1.0, 4.0, 1.5, 2.0, 3.5, 1.5, 3.0, 4.0, 3.5, 1.0]

$\mu = 2.5,\ \sigma = 1.16$

$x = 3.5$

$$pdf(3.5, 2.5, 1.16) = \frac{1}{1.16\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{3.5-2.5}{1.16}\right)^2} = 0.237$$

Study Hours of Passed = [6, 5, 6, 7, 3, 6]

$\mu = 5.5,\ \sigma = 1.378$

$x = 5$

$$pdf(5, 5.5, 1.378) = \frac{1}{1.378\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{5-5.5}{1.378}\right)^2} = 0.271$$

Study Hours (All) = [6, 5, 2, 6, 7, 4, 3, 2, 6, 4]

$\mu = 4.5, \sigma = 1.69$

$x = 5$

$$pdf(5, 4.5, 1.69) = \frac{1}{1.69\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{5-4.5}{1.69}\right)^2} = 0.226$$

$$P(Pass \mid 3.5,\ 5) = \frac{P(3.5 \mid Pass)P(5 \mid Pass)P(Pass)}{P(3.5, 5)} = \frac{(0.292)(0.271)(0.6)}{(0.237)(0.226)} = \frac{0.0474792}{0.053562} = 0.88643441$$

6. P(Fail | 3.5, 5)

Math Grades of Failed = [1.5, 1.5, 4.0, 1.0]

$\mu = 2, \sigma = 1.354$

$x = 3.5$

$$pdf(3.5, 2, 1.354) = \frac{1}{1.354\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{3.5-2}{1.354}\right)^2} = 0.16$$

Study Hours of Failed = [2, 4, 2, 4]

$\mu = 3, \sigma = 1.155$

$x = 5$

$$pdf(5, 3, 1.155) = \frac{1}{1.155\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{5-3}{1.155}\right)^2} = 0.077$$

$$P(Fail \mid 3.5,\ 5) = \frac{P(3.5 \mid Fail)P(5 \mid Fail)P(Fail)}{P(3.5, 5)} = \frac{(0.16)(0.077)(0.4)}{(0.237)(0.226)} = \frac{0.004928}{0.053562} = 0.09200553$$