CSC 39002 – Machine Learning – HW W03 (30 pts) **Name** Key

Problem 1. (6 pts) A training set and a testing set for a regression task are provided below. The datasets have two features, $x^{(1)}$ and $x^{(2)}$, and one continuous label, y. You are asked to calculate the testing r^2 for this problem using two models: simple linear regression, and K-nearest neighbors regression. Recall the following formulas relating to the calculation of r^2 for a regression task:

- $SST = \sum(y_i - \bar{y})^2$
- $SSE = \sum \hat{e}_i^2$
- $\hat{e}_i = y_i - \hat{y}_i$
- $r^2 = 1 - \frac{SSE}{SST}$

Training Set

$x^{(1)}$	3	7	2	8	1
$x^{(2)}$	3	4	1	3	4
y	6.77	20.25	9.83	22.37	-1.42

Testing Set

$x^{(1)}$	8	1	5
$x^{(2)}$	2	7	6
y	25.83	0.30	8.86

Round all answers to 2 decimal places on this problem.

a) **Find SST.** Begin by finding SST for the ~~training~~ testing data.

$\bar{y} = \frac{1}{3}(25.83 + 0.30 + 8.86) = 11.6633$

$SST = (25.83 - 11.6633)^2 + (0.30 - 11.6633)^2 + (8.86 - 11.6633)^2$

$= \boxed{337.68}$

b) **Simple Linear Regression.** Applying a simple linear regression model to the training data produces the model:

$$\hat{y} = 4.848 + 3.239x^{(1)} - 2.298x^{(2)}$$

Use this model to complete the following table for the testing observations:

$x^{(1)}$	8	1	5
$x^{(2)}$	2	7	6
y	25.83	0.3	8.86
$\hat{y}$	26.164	-7.999	7.255
$\hat{e}$	-0.334	8.299	1.605

Now calculate testing **SSE** and r^2 for the simple linear regression model.

$SSE = 0.334^2 + 8.299^2 + 1.605^2 = 71.56$

$r^2 = 1 - \frac{71.56}{337.68} = \boxed{0.7881}$

c) **2-Nearest Neighbors Regression.** In the table below, there is one row for each testing observation and one column for each training observation. Calculate the distance between each training observation and each testing observation and write the result in the appropriate cell in the table. In the last two columns, find the predicted label values for each testing observation according to a 2-nearest neighbors regression model. Also calculate the residual (error term) for each testing observation.

			y	6.77	20.25	9.83	22.37	-1.42		
			$x^{(1)}$	3	7	2	8	1		
			$x^{(2)}$	3	4	1	3	4		
y	$x^{(1)}$	$x^{(2)}$							$\hat{y}$	$\hat{e}$
25.83	8	2		5.10	2.24	6.08	1	7.28	21.31	4.52
0.30	1	7		4.47	6.71	6.08	8.06	3	2.675	-2.375
8.86	5	6		3.61	2.83	5.83	4.24	4.47	13.51	-4.65

Now calculate testing **SSE** and r^2 for the 2-nearest neighbors regression model.

$$SSE = 4.52^2 + 2.375^2 + 4.65^2 = \boxed{47.69}$$

$$r^2 = 1 - \frac{47.69}{337.68} = \boxed{0.8588}$$

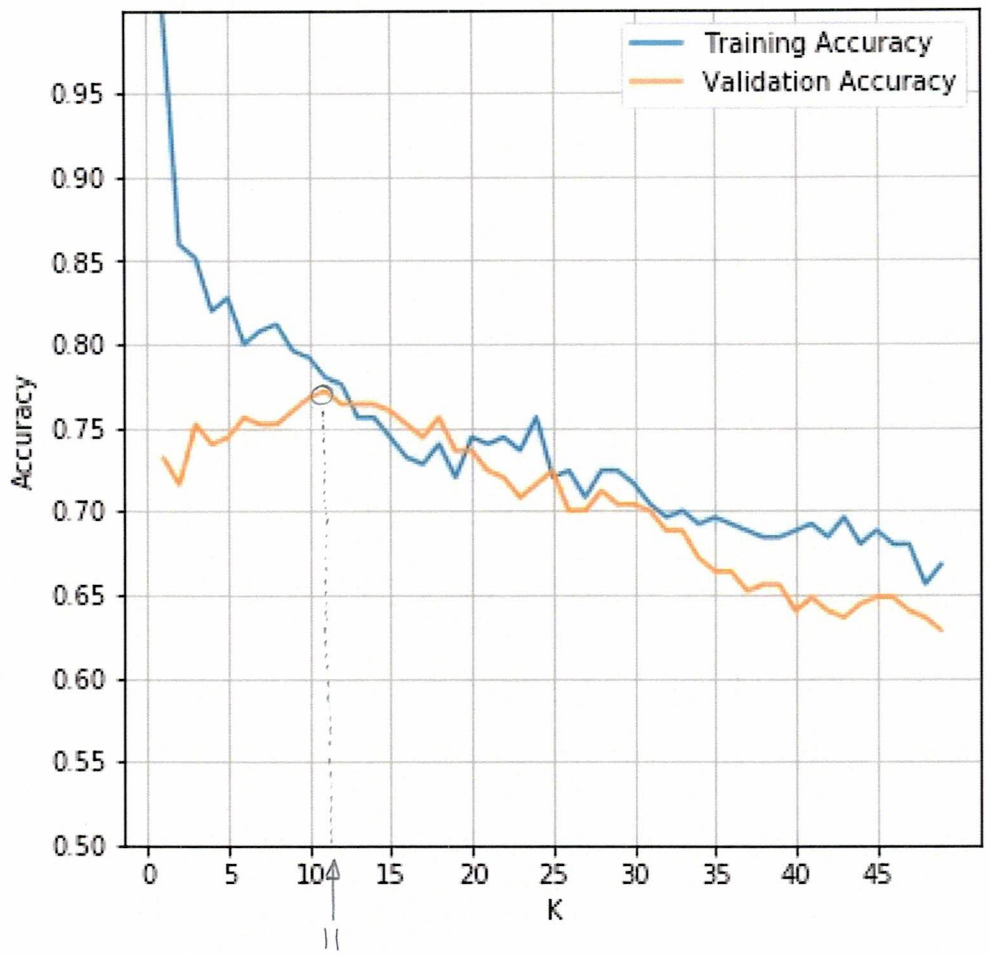
Problem 2. (3 pts) The confusion matrix for a testing set in a classification problem with three classes is provided below. Find the precision and recall for each class, as well as the overall accuracy of the model. **Round to three decimal places.**

	Class 0	**Class 1**	**Class 2**	
Class 0	20	4	8	32
Class 1	4	4	4	12
Class 2	2	2	16	20
	26	10	28	64

	Precision	**Recall**
Class 0	0.769	0.625
Class 1	0.400	0.333
Class 2	0.571	0.800

Accuracy: 0.625

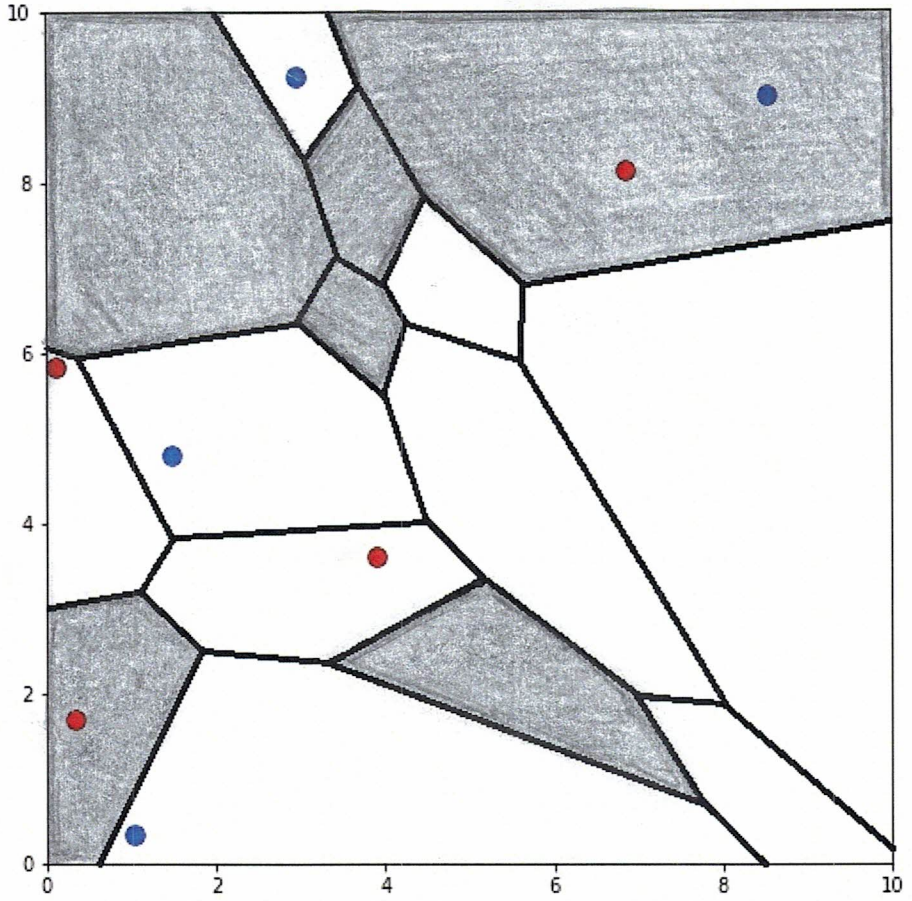
Problem 3. (3 pts) The graph below shows how the training and validation accuracy for a K-nearest neighbors classification problem vary for different values of *K*. Determine the value of *K* that would be the best to use for this task. Explain your answer.



K=11

This is the value of K that results in the highest validation accuracy. This suggests that K=11 would produce the model that would generalize to new data the best.

Problem 4. (6 pts) A scatter plot consisting of eight points is shown below. Each point is labelled as one of two classes: "red" or "blue". Assume a KNN classification algorithm is trained on this dataset with K=3. Determine how points in each of the displayed regions would be classified using this algorithm. Shade in any region that would be classified as the "blue" class, and leave blank any region that would be classified as the "red" class. A compass right be useful for this problem.



Problem 5. (6 pts) A training set for a regression task is provided below. The dataset has three features and one continuous label, y. You are asked to score two potential models using the **lasso** cost function.

Round all answers to two decimal places on this problem.

$x^{(1)}$	8	3	6	3	5
$x^{(2)}$	8	1	3	5	6
$x^{(3)}$	6	7	6	2	3
y	-6.1	8.8	2.7	-8.4	-1.4

a) Generate predictions according to each of the models below. Also calculate the residuals.

Model 1: $\hat{y} = 0.2 + 0.1x^{(1)} - 1.5x^{(2)} + 1.2x^{(3)}$

$\hat{y}$	-3.8	7.4	3.5	-4.6	-4.7
$\hat{e}$	-2.3	1.4	-0.8	-3.8	3.3

Model 2: $\hat{y} = 0.9 + 0.3x^{(1)} - 1.9x^{(2)} + 1.1x^{(3)}$

$\hat{y}$	-5.3	7.6	3.6	-5.5	-5.7
$\hat{e}$	-0.8	1.2	-0.9	-2.9	4.3

b) Score each of the models using the lasso cost function with $\alpha = 2.5$. Assuming that there are n observations and p features, the lass0 cost function is given by:

$$COST = \frac{SSE}{n} + \alpha \sum_{i=1}^{p} |\beta_i| = \frac{SSE}{n} + \alpha \left(|\beta_2| + |\beta_2| + \cdots + |\beta_p| \right)$$

Model 1 Cost:

$$\text{Cost} = \frac{1}{5}(2.3^2 + 1.4^2 + 0.8^2 + 3.8^2 + 3.3^2) + 2.5(0.1 + 1.5 + 1.2)$$

$$= \boxed{13.644}$$

Model 2 Cost:

$$\text{Cost} = \frac{1}{5}(0.8^2 + 1.2^2 + 0.9^2 + 2.9^2 + 4.3^2) + 2.5(0.3 + 1.9 + 1.1)$$

$$= \boxed{14.208}$$

c) Which model has a better score?

Model 1

Problem 6. (6 pts) A training set for a classification task is provided below. The dataset has two features and one categorical label, y, which has two possible classes: "red" and "blue". You are asked to score two logistic regression models. **Round all answers to three decimal places on this problem.**

$x^{(1)}$	1	5	5	6	7
$x^{(2)}$	4	1	7	3	8
y	blue	blue	red	red	blue

a) Two logistic regression models are provided below. In these models, $\hat{p}$ is an estimate for the probability that an observation falls into the red class. Let $\hat{\pi} = \hat{p}$ for red observations and let $\hat{\pi} = 1 - \hat{p}$ for blue observations. For each model, find $\hat{p}$ and $\hat{\pi}$ for each observation.

Model 1: $\hat{p} = 1/\left[1 + \exp\left(-2 + 0.3x^{(1)} + 0.1x^{(2)}\right)\right]$

$\hat{p}$	0.786	0.599	0.450	0.475	0.289
$\hat{\pi}$	0.214	0.401	0.450	0.475	0.711

Model 2: $\hat{p} = 1/\left[1 + \exp\left(-4 + 0.2x^{(1)} + 0.4x^{(2)}\right)\right]$

$\hat{p}$	0.900	0.931	0.550	0.832	0.354
$\hat{\pi}$	0.100	0.069	0.550	0.832	0.646

b) Calculating the log-likelihood score for each model.

Model 1 Log-Likelihood:

$$\ln(L) = \ln(0.214) + \ln(0.401) + \ln(0.450) + \ln(0.475) + \ln(0.711)$$
$$= \boxed{-4.340}$$

Model 2 Log-Likelihood:

$$\ln(L) = \ln(0.100) + \ln(0.069) + \ln(0.550) + \ln(0.832) + \ln(0.646)$$
$$= \boxed{-6.195}$$

c) Which model has a better score?

Model 1