Problem 1. (10 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, and show your work.**

$x^{(1)}$	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 negative log-likelihood score for each model.

Model 1 Negative Log-Likelihood:

$$NLL = -\left[\ln(0.214) + \ln(0.401) + \ln(0.450) + \ln(0.475) + \ln(0.711)\right]$$

$$= \boxed{4.340}$$

Model 2 Negative Log-Likelihood:

$$NLL = -\left[\ln(0.100) + \ln(0.069) + \ln(0.550) + \ln(0.832) + \ln(0.646)\right]$$

$$= \boxed{6.195}$$

c) Which model has a better score?

$$\boxed{\text{Model 1}}$$

Problem 2 (6 pts). The confusion matrix for a test 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:

Problem 3 (2 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.
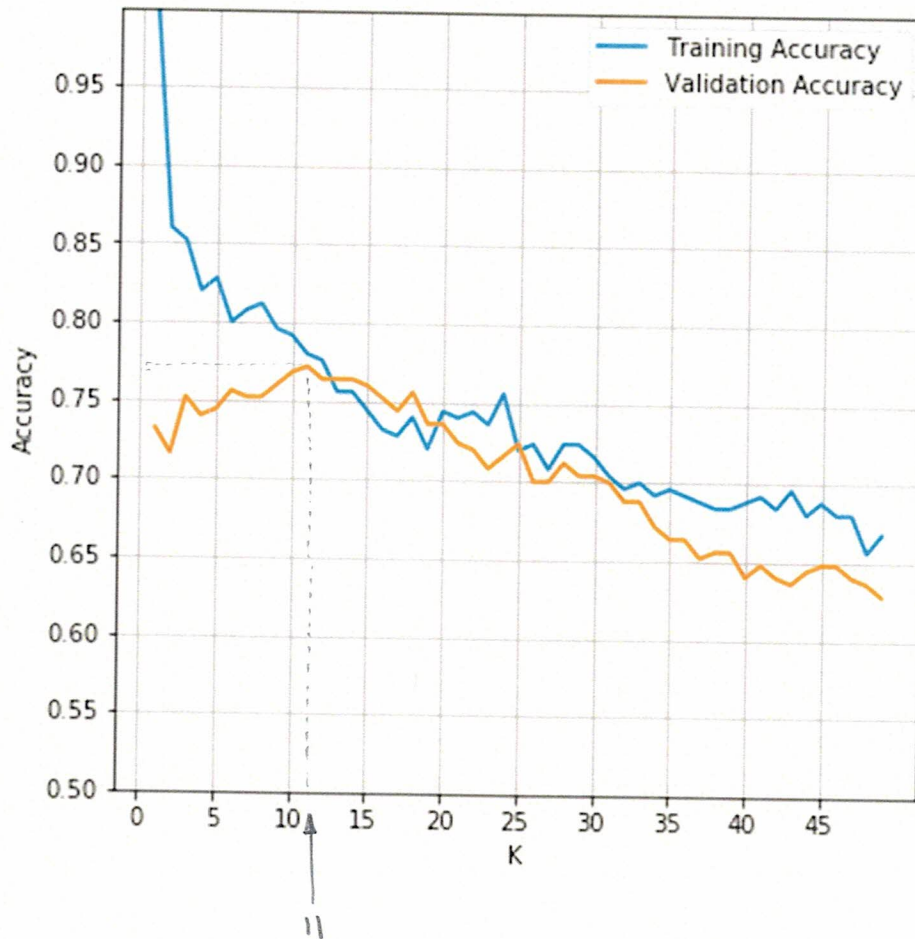
$K = 11$

This is where validation accuracy is maximized.

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.
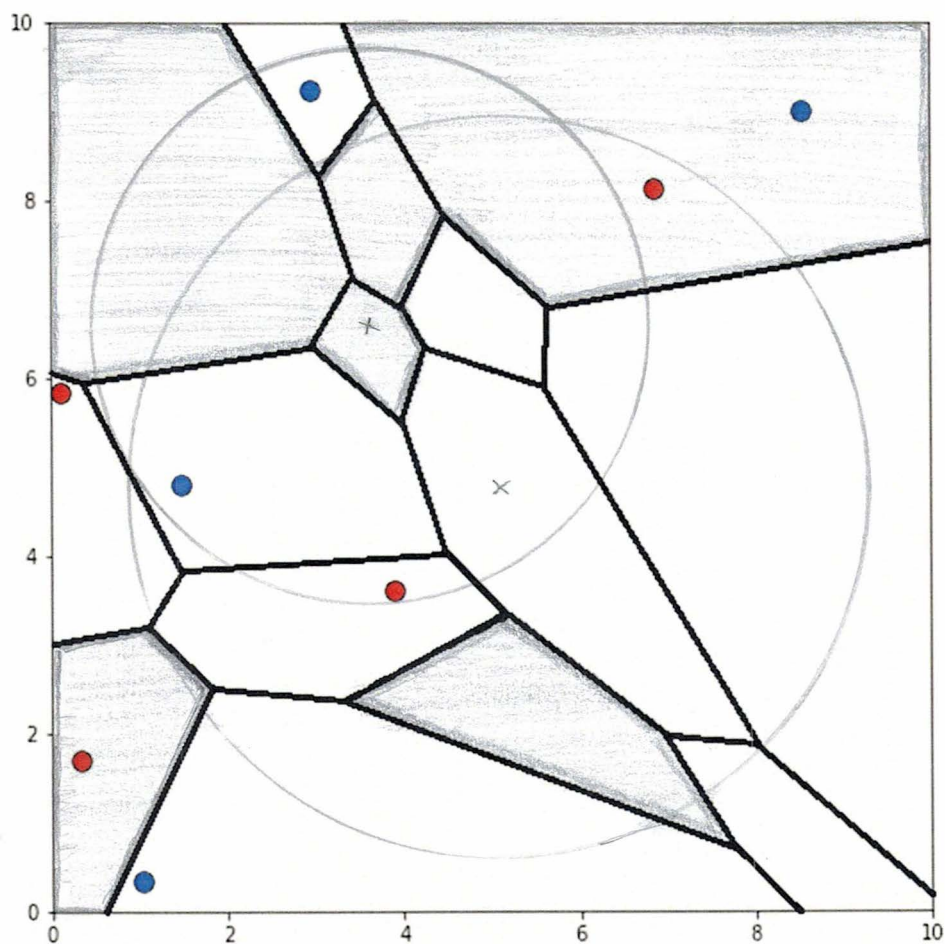
Problem 5. (12 pts) Write **True** or **False** next to each of the following statements.

This problem will be graded as follows: +1 points for each correct answer, -1 points for each incorrect answer, and 0 points for each blank answer. So, for instance, one correct answer and one incorrect answer will cancel each other out. You cannot get less than zero points on this problem, even if every answer is incorrect.

a) It is important to scale features when constructing a ridge regression model.

True

b) It is important to scale features when constructing a basic linear regression model.

False

c) It is important to scale features when constructing a KNN model.

True

d) The output of a logistic regression model is an estimate of the probability that the provided observation belongs to a particular class.

True

e) The advantage of the lasso algorithm over standard linear regression is that the lasso algorithm will generally produce a model with a lower training SSE than linear regression.

False

f) The lasso algorithm is a regularized version of least squares regression that tries to avoid overfitting by applying a penalty to models based on the size of their coefficients.

True

g) We can avoid encountering ties when using the KNN algorithm for any given classification task by selecting K to be an odd number.

False

h) The coefficients in a logistic regression model are trained by minimizing the sum of squared errors objective function.

False

i) We use negative log-likelihood rather than likelihood when score a logistic regression model because negative log-likelihood is less affected by rounding issues.

True

j) Training a regression model using the mean-squared error loss function will tend to produce smaller coefficient estimates than training with sum of squared error loss.

False

k) Increasing the hyper-parameter α in a ridge regression model will typically make it less likely that the model will overfit.

True

l) Increasing the hyper-parameter K in a KNN model will typically make it less likely that the model will overfit.

True