

# The University of Texas at Austin

## IN-CLASS WORK 2

*M339G Predictive Analytics*

February 26, 2026

---

### K-NEAREST NEIGHBORS.

**Problem 2.1.** (10 points) Explain why standardization is an appropriate preliminary step before we employ the  $K$ -nearest neighbors procedure.

**Solution.** Solutions will vary. The salient point of any response which is to earn credit must be that  $K$ -nearest neighbors employs a distance which is heavily affected by the natural scale of various predictors.

**Problem 2.2.** (10 points) You are using  $K$ -nearest neighbors in a classification problem with  $X$  as the explanatory variable and  $Y$  as the response. Here are the observed values:

x	1	2	5.5	6.5	9	13
y	1	2	1	1	2	2

Using  $K = 3$ , figure out the misclassification error.

**Solution.** With the neighbourhood of size 3, we get the following predictions:

x	1	2	5.5	6.5	9	13
$\hat{y}$	1	1	1	1	1	2

We did fine: the misclassification error rate is  $1/3$ .

**Problem 2.3.** (10 points) You are using  $K$ -nearest neighbors in a classification problem with  $X = (X_1, X_2)$  as predictors and  $Y$  as the response. Here are the observed values:

$x_1$	1	1	2	2
$x_2$	1	4	2	-1
y	1	2	1	2

Using  $K = 3$ , figure out how the above points would be classified and the misclassification error. Then, state how you would classify point  $(2, 4)$ .

*Hint: Draw a picture in the plane of  $(x_{i1}, x_{i2})$  for  $i = 1, 2, 3, 4$ .*

**Solution.** With the neighbourhood of size 3, we get the following predictions:

$x_1$	1	1	2	2
$x_2$	1	4	2	-1
y	1	1	1	1

We did fine: the misclassification error rate is  $1/2$ . The new point would be classified as 1.