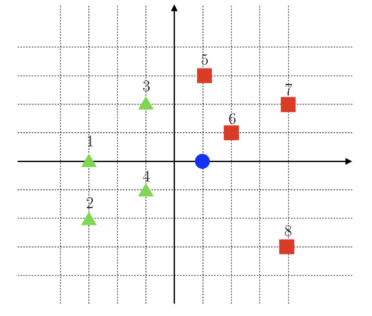
Problem 1. (kNN) 4 points
Classify the new point (circle) using a kNN classifier. Choose the closest $k = 3$ neighbors among the data points 1 to 8 and their class label (△ or □) for different distance metrics.



	Closest $k = 3$ points	Label
L1 (Manhattan) distance	1, 2, 3, 4, 5, 6, 7, 8	□, △
L2 (Euclidean) distance	1, 2, 3, 4, 5, 6, 7, 8	□, △

Solution:

	Closest $k = 3$ points	Label
L1 (Manhattan) distance	4,5,6	□
L2 (Euclidean) distance	3,4,6	△

Problem 2. (Naive Bayes) 3 points
We aim to predict the possibility of success based on data point $x = (x_1, \dots, x_d)$ with features $x_i \in \mathbb{R}$, $i = 1, \dots, d$. Success corresponds to class 1 and failure corresponds to class 0. Let $P(1)$ denote the probability of success determined based on a training set.

1. (2 points) Suppose you are given a test point $x^t = (x_1^t, \dots, x_d^t)$ and $f_{x_i^t|1}$ is the conditional probability density of feature x_i given success. Give the expression of the probability that the test point is a success, by filling in the blanks below and using feature independence with the Naive Bayes assumption.

   Solution:

$$
\begin{aligned}
P\left(1 \mid x^t\right) =& \frac{f_{x^t|1}(x^t)P(1)}{f_{x^t}(x^t)} \\
\propto& \prod_{i=1}^{d} f_{x_i^t|1}(x_i^t)P(1).
\end{aligned} \tag{0.1}
$$

   Comment: The problem statement contains a typo, specifically with '$f_{x^t|1}$' in the original version. Consequently, students write (0.1) will still receive full points.

2. (1 points) Suppose the number of features is $d = 1$ and the test data point is $x^t = 4$. Assume the following is known to you: $f_{x^t|1}(4) = 0.17$, $f_{x^t|0}(4) = 0.24$ and $P(1) = 0.6$. Using the Naive Bayes approach to determine whether the test point $x^t = 4$ is classified as a success or not.

   Solution:

i Compute

$$
\begin{aligned}
P\left(1 \mid x^t\right) =& \frac{f_{x^t|1}(x^t)P(1)}{f_{x^t}(x^t)} \\
\propto& f_{x_1^t|1}(x_1^t)P(1) \\
\propto& f_{4|1}(x_1^t) * 0.6 \\
\propto& 0.102
\end{aligned}
$$

ii Compute

$$
\begin{aligned}
P\left(0 \mid x^t\right) =& \frac{f_{x^t|0}(x^t)P(0)}{f_{x^t}(x^t)} \\
\propto& f_{x_1^t|0}(x_1^t)P(0) \\
\propto& f_{4|0}(x_1^t) * 0.4 \\
\propto& 0.096
\end{aligned}
$$

iii Since $P\left(1 \mid x^t\right) > P\left(0 \mid x^t\right)$ the test point is classified as a success.

Problem 3. (Label axes (i.e. x_1, x_2)) 3 points

Consider the following classification problem with data points $x^i \in \mathbb{R}^2$ and corresponding labels $y^i \in \{0\,(\text{square}), 1\,(\text{triangle})\}$. The training data is depicted in Figure 1 below. Your friend suggests calculating $\hat{z}^i$ as $\hat{z}^i := (x_1^i)^2 + (x_2^i)^2 - R^2$, where R represents the radius of the circle in the graph. Given $\hat{z}^i$, you assign x^i to label 1 if $\hat{z}^i \geq 0$ and label 0 if $\hat{z}^i < 0$.

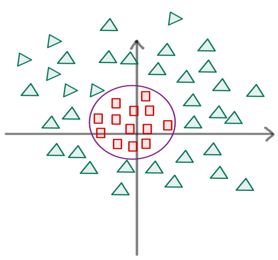


Figure 1: Classification problem with training data

1. (3 points) Based on your friend's suggestion, what is the feature function $\Phi(x_1^i, x_2^i) : \mathbb{R}^2 \to \mathbb{R}$? What are the values of the weight $w \in \mathbb{R}$ and the bias $b \in \mathbb{R}$?

   Solution: Feature function is $\Phi(x_1^i, x_2^i) = \left(x_1^i\right)^2 + \left(x_2^i\right)^2$, and $w = 1$ and $b = -R^2$.