ARTIFICIAL INTELLIGENCE (E016350)
Prof. Aleksandra Pizurica
AY 2024/2025

Nonparametric ML models

1. Select all that apply about k Nearest Neighbors (kNN) in the following options:
   Assume a point can be its own neighbor.
   - ☐ k-NN works great with a small amount of data, but struggles when the amount of data becomes large.
   - ☐ k-NN is sensitive to outliers; therefore, in general we decrease k to avoid overfitting.
   - ☐ k-NN can only be applied to classification problems, but it cannot be used to solve regression problems.
   - ☐ We can always achieve zero training error (perfect classification) with k-NN, but it may not generalize well in testing.

2. Suppose a 7-nearest-neighbors regression search returns $\{7, 6, 8, 4, 7, 11, 100\}$ as the 7 nearest y values for a given x value. What is the value of $\hat{y}$ that minimizes the L_1 loss function on this data? There is a common name in statistics for this value as a function of the y values; what is it? Answer the same two questions for the L_2 loss function.

3. Figure 1 shows how a circle at the origin can be linearly separated by mapping from the features (x_1, x_2) to the two dimensions (x_1^2, x_2^2). But what if the circle is not located at the origin? What if it is an ellipse, not a circle? The general equation for a circle (and hence the decision boundary) is $(x_1 - a)^2 + (x_2 - b)^2 - r^2 = 0$, and the general equation for an ellipse is $c(x_1 - a)^2 + d(x_2 - b)^2 - 1 = 0$.

    1. Expand out the equation for the circle and show what the weights w_i would be for the decision boundary in the four-dimensional feature space (x_1, x_2, x_1^2, x_2^2). Explain why this means that any circle is linearly separable in this space.
    2. Do the same for ellipses in the five-dimensional feature space $(x_1, x_2, x_1^2, x_2^2, x_1 x_2)$.

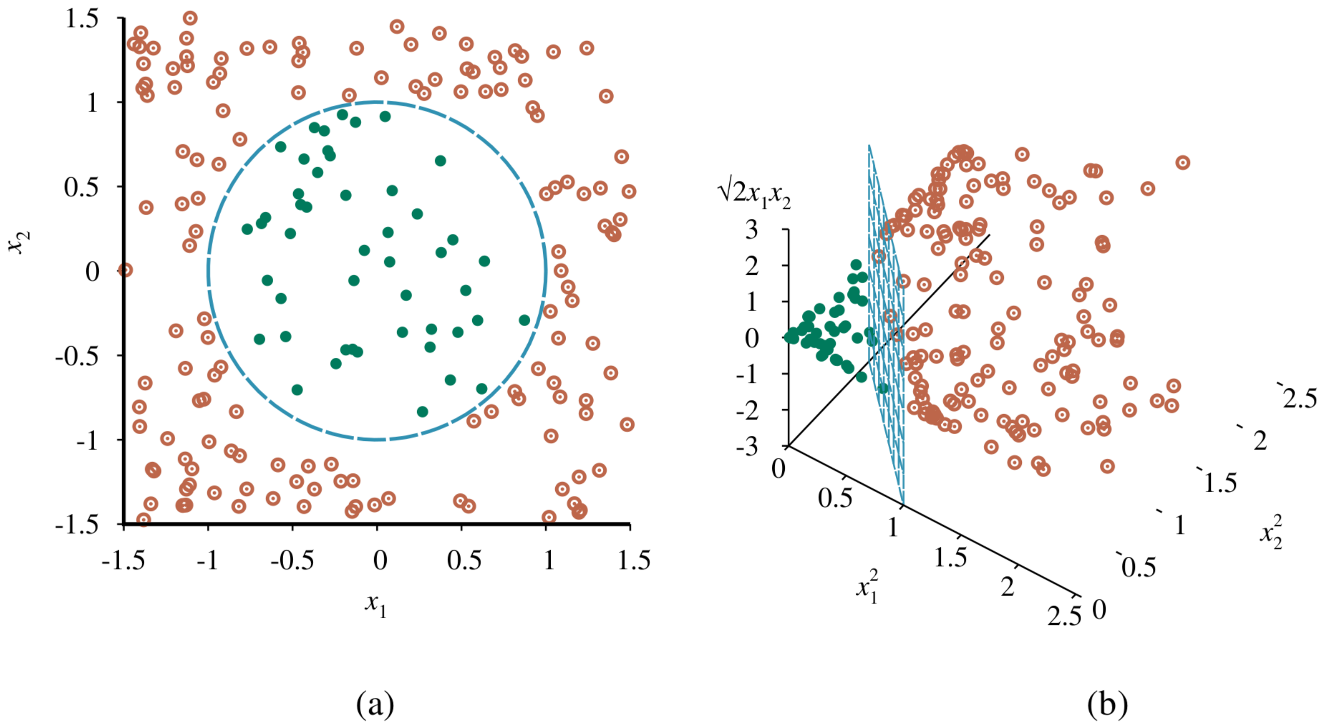

Figure 1: (a) A two-dimensional training set with positive examples as green filled circles and negative examples as orange open circles. The true decision boundary, $x_1^2 + x_2^2 \leq 1$, is also shown. (b) The same data after mapping into a three-dimensional input space $(x_1^2, x_2^2, \sqrt{2}x_1x_2)$. The circular decision boundary in (a) becomes a linear decision boundary in three dimensions. Figure from *Artificial Intelligence: A Modern Approach, 4th US ed.*, Russel and Norvig.

4. Construct a support vector machine that computes the XOR function. Use values of +1 and –1 (instead of 1 and 0) for both inputs and outputs, so that an example looks like $([-1, 1], 1)$ or $([-1, -1], -1)$. Map the input $[x_1, x_2]$ into a space consisting of x_1 and $x_1 x_2$. Draw the four input points in this space, and the maximal margin separator. What is the margin? Now draw the separating line back in the original Euclidean input space.