GHENT UNIVERSITY

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

Solutions: 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.

Solution:

- ■ k-NN works great with a small amount of data, but struggles when the amount of data becomes large. *(True because k-NN is slow and imposes high memory requirements.)*
- ☐ k-NN is sensitive to outliers; therefore, in general we decrease k to avoid overfitting. *(It's the opposite: we increase k to avoid overfitting)*
- ☐ k-NN can only be applied to classification problems, but it cannot be used to solve regression problems.
  (Can yield regression by averaging the data in the same neighbourhood)
- ■ We can always achieve zero training error (perfect classification) with k-NN, but it may not generalize well in testing. *(By setting $k = 1$)*

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.

Solution:

- The L_1 loss is minimized by the median, in this case 7.

  Detail: Suppose we have an odd number $2n+1$ of elements $y_{-n} < \ldots < y_0 < \ldots < y_n$. For $n = 0$, $\hat{y} = y_0$ is the median and it minimizes the loss. Then, observe that the L_1 loss for $n+1$ is

  $$\frac{1}{2n+3} \sum_{i=-(n+1)}^{n+1} |\hat{y} - y_i| = \frac{1}{2n+3} \left(|\hat{y} - y_{n+1}| + \left|\hat{y} - y_{-(n+1)}\right| \right) + \frac{1}{2n+3} \sum_{i=-n}^{n} |\hat{y} - y_i|$$

  The first term equals $|y_{n+1} - y_{-(n+1)}|$ whenever $y_{n+1} \leq \hat{y} \leq y_{-(n+1)}$, e.g. for $\hat{y} = y_0$, and is strictly larger otherwise. By inductive hypothesis the second term is also minimized by $\hat{y} = y_0$, the median.

- The L_2 loss is minimized by the mean, in this case $\frac{143}{7} \approx 20.4$.

  Detail: Note that the L_2 loss of $\hat{y}$ given data $y_1, \ldots, y_n$ is

  $$\frac{1}{n} \sum_i (\hat{y} - y_i)^2.$$

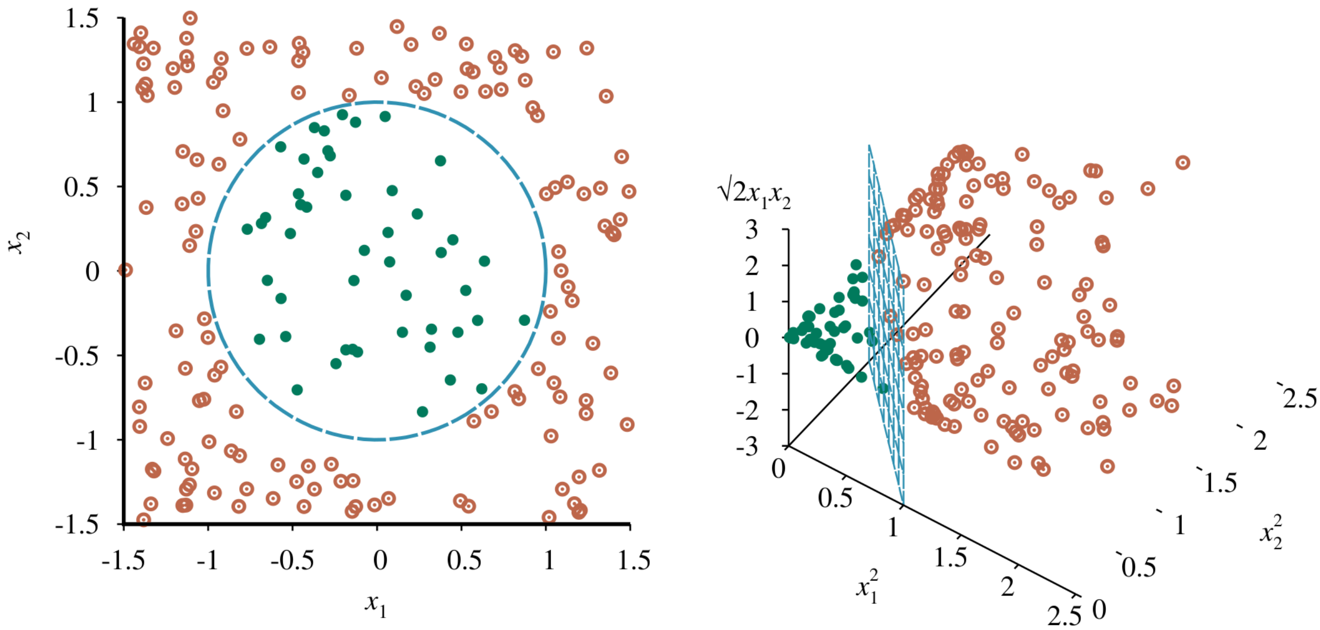
  This loss is differentiable so we can find critical points:

  $$0 = \frac{2}{n} \sum_i (\hat{y} - y_i),$$

  or $\hat{y} = (1/n) \sum_i y_i$. Taking the second derivative we see this is the unique local minimum, and thus the global minimum as the loss tends to infinite when $\hat{y}$ tends to either infinity.

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 $(x_1, x_2, x_1^2, x_2^2, x_1 x_2)$.



(a) (b)

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 $(x_1^2, x_2^2, \sqrt{2}x_1 x_2)$. 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.

Solution:

1. The circle equation expands into five terms

$$0 = x_1^2 + x_2^2 - 2ax_1 - 2bx_2 + (a^2 + b^2 - r^2)$$

corresponding to the weights $\mathbf{w} = [-2a, -2b, 1, 1]^\top$ and and the bias term $a^2 + b^2 - r^2$. This shows that a circular boundary is linear in this feature space, allowing linear separability. In fact, the three features $x_1, x_2, x_1^2 + x_2^2$ suffice.

2. The (axis-aligned) ellipse equation expands into six terms

$$0 = cx_1^2 + dx_2^2 - 2acx_1 - 2bdx_2 + (a^2c + b^2d - 1)$$

corresponding to the weights $\mathbf{w} = [-2ac, -2bd, c, d, 0]^\top$ and the bias term $a^2c + b^2d - 1$. This shows that an elliptical boundary is linear in this feature space, allowing linear separability. In fact, the four features x_1, x_2, x_1^2, x_2^2 suffice for any axis-aligned ellipse.

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_1x_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.

Solution:
The examples map from $[x_1, x_2]$ to $[x_1, x_1x_2]$ coordinates as follows
$[-1, -1]$ (negative) maps to $[-1, +1]$,
$[-1, +1]$ (positive) maps to $[-1, -1]$,
$[+1, -1]$ (positive) maps to $[+1, -1]$,
$[+1, +1]$ (negative) maps to $[+1, +1]$.
Thus the positive examples have $x_1x_2 = -1$ and the negative examples have $x_1x_2 = +1$. The maximum margin separator is the line $x_1x_2 = 0$, with a margin of 1. The separator corresponds to the $x_1 = 0$ and $x_2 = 0$ axes in the original spaces; this can be thought of as the limit of a hyperbolic separator with two branches.