P04 - KNN Linear and Regression

Luis Sa-Couto[1] and Andreas Wichert[2]

INESC-ID, Instituto Superior Tecnico, Universidade de Lisboa
{luis.sa.couto,andreas.wichert}@tecnico.ulisboa.pt

1 Nearest Neighbour

To illustrate this idea let us describe a simple example task. Assuming that 1 means $True$ and 0 means $False$, consider the following features and class:

- X_1: "Weight (Kg)"
- X_2: "Height (Cm)"
- $Class$: "NBA (National Basketball Association) Player"

Consider as well that you are given the following training set:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$	$\mathbf{x}^{(7)}$	$\mathbf{x}^{(8)}$	$\mathbf{x}^{(9)}$	$\mathbf{x}^{(10)}$
X_1	170	80	90	60	50	70	90	100	110	80
X_2	160	220	200	160	150	190	170	180	178	210
$Class$	0	1	1	0	0	1	0	0	0	0

1) a) Using the same data from the example above, use a k nearest neighbors classifier to classify vector $\mathbf{x} = \begin{bmatrix} 100 & 210 \end{bmatrix}^T$ with $k = 1$, $k = 3$ and $k = 5$.

---

b) Redo the exercise computing the distance with the $l1$ norm $\|\mathbf{x}\|_1 = \sum_i |x_i|$.

---

c) Again, repeate the exercise computing the distance with the infinity norm $\|\mathbf{x}\|_\infty = max_i\,|x_i|$.

---

2) Assuming that 1 means $True$ and 0 means $False$, consider the following features and class:

- X_1: "Fast processing"
- X_2: "Decent Battery"
- X_3: "Good Camera"
- X_4: "Good Look and Feel"
- X_5: "Easiness of Use"
- $Class$: "iPhone"

You are given the following training set:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$	$\mathbf{x}^{(6)}$	$\mathbf{x}^{(7)}$
X_1	1	1	0	0	1	0	0
X_2	1	1	1	0	0	0	0
X_3	0	1	1	0	1	1	0
X_4	1	0	1	1	1	0	0
X_5	0	0	0	1	1	0	1
$Class$	1	0	0	0	1	1	1

Use a k nearest neighbors classifier based on the Hamming distance to classify vector $\begin{bmatrix}1\ 1\ 1\ 1\ 1\end{bmatrix}^T$ with $k = 1$ and $k = 3$.

3) Consider the following data where a few preprocessed restaurant reviews (without stopwords) are classified as positive (1) or negative (0).

$Sentence$	$Class$
$\{"Great", "place", "go", "with", "friends"\}$	1
$\{"Food", "amazing"\}$	1
$\{"What", "terrible", "experience", "no", "words"\}$	0
$\{"Waiting", "time", "too", "long"\}$	0
$\{"Terrible", "place", "for", "family", "dinner"\}$	0

Consider as well the Jaccard similarity between two sets:

$$Jaccard_{sim}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Using the similarity measure, compute the k nearest neighbor output for input $\{"Terrible", "food", "overall", "lousy", "dinner"\}$ using $k = 1$ and $k = 3$.

2 Closed form learning

1) Consider the following training data:

$$\left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\1\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}2\\1\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}1\\3\end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix}3\\3\end{pmatrix}\right\}$$

$$\left\{t^{(1)} = 1.4, t^{(2)} = 0.5, t^{(3)} = 2, t^{(4)} = 2.5\right\}$$

a) Find the closed form solution for a linear regression that minimizes the sum of squared errors on the training data..

b) Predict the target value for $x_{query} = \begin{pmatrix}2\ 3\end{pmatrix}^T$.

c) Sketch the predicted hyperplane along which the linear regression predicts points will fall.

d) Compute the mean squared error produced by the the linear regression.

2) Consider the following training data:

$$\left\{\mathbf{x}^{(1)} = (-2.0)\,, \mathbf{x}^{(2)} = (-1.0)\,, \mathbf{x}^{(3)} = (0.0)\,, \mathbf{x}^{(4)} = (2.0)\right\}$$

$$\left\{t^{(1)} = 2.0, t^{(2)} = 3.0, t^{(3)} = 1.0, t^{(4)} = -1.0\right\}$$

a) Find the closed form solution for a linear regression that minimizes the sum of squared errors on the training data..

b) Predict the target value for $x_{query} = (1)^T$.

c) Sketch the predicted hyperplane along which the linear regression predicts points will fall.

d) Compute the mean squared error produced by the the linear regression.

3) Consider the following training data:

$$\left\{\mathbf{x}^{(1)} = \begin{pmatrix}1\\1\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}2\\1\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}1\\3\end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix}3\\3\end{pmatrix}\right\}$$

$$\left\{t^{(1)} = 1, t^{(2)} = 1, t^{(3)} = 0, t^{(4)} = 0\right\}$$

a) Find the closed form solution for a linear regression that minimizes the sum of squared errors on the training data..

b) Use your linear regression to classify $x_{query} = \begin{pmatrix}2 & 2.5\end{pmatrix}^T$, assuming a threshold similarity of 0.5.

4) Consider the following training data:

$$\left\{\mathbf{x}^{(1)} = (-2.0)\,, \mathbf{x}^{(2)} = (-1.0)\,, \mathbf{x}^{(3)} = (0.0)\,, \mathbf{x}^{(4)} = (2.0)\right\}$$

$$\left\{t^{(1)} = 1, t^{(2)} = 0, t^{(3)} = 0, t^{(4)} = 0\right\}$$

a) Find the closed form solution for a linear regression that minimizes the sum of squared errors on the training data..

b) Use your linear regression to classify $x_{query} = (-0.3)^T$, assuming a threshold similarity of 0.15.

5) Consider the following training data:

$$\mathbf{x}^{(1)} = \begin{pmatrix}-0.95\\0.62\end{pmatrix}, \mathbf{x}^{(2)} = \begin{pmatrix}0.63\\0.31\end{pmatrix}, \mathbf{x}^{(3)} = \begin{pmatrix}-0.12\\-0.21\end{pmatrix}, \mathbf{x}^{(4)} = \begin{pmatrix}-0.24\\-0.5\end{pmatrix},$$

$$\mathbf{x}^{(5)} = \begin{pmatrix}0.07\\-0.42\end{pmatrix}, \mathbf{x}^{(6)} = \begin{pmatrix}0.03\\0.91\end{pmatrix}, \mathbf{x}^{(7)} = \begin{pmatrix}0.05\\0.09\end{pmatrix}, \mathbf{x}^{(8)} = \begin{pmatrix}-0.83\\0.22\end{pmatrix}$$

$$\left\{t^{(1)} = 0, t^{(2)} = 0, t^{(3)} = 1, t^{(4)} = 0, t^{(5)} = 1, t^{(6)} = 0, t^{(7)} = 1, t^{(8)} = 0\right\}$$

a) Plot the data points and try to choose a non-linear transformation to apply.

b) Adopt the non-linear transform you chose in a) and find the closed form solution.

c) Sketch the predicted surface along which the predictions will fall.

6) Consider the following training data:

$$\left\{\mathbf{x}^{(1)} = (3), \mathbf{x}^{(2)} = (4), \mathbf{x}^{(3)} = (6), \mathbf{x}^{(4)} = (10), \mathbf{x}^{(5)} = (12)\right\}$$

$$\left\{t^{(1)} = 1.5, t^{(2)} = 9.3, t^{(3)} = 23.4, t^{(4)} = 45.8, t^{(5)} = 60.1\right\}$$

a) Adopt a logarithmic feature transformation $\phi(x_1) = \log(x_1)$ and find the closed form solution for this non-linear regression that minimizes the sum of squared errors on the training data.

b) Repeat the exercise above for a quadratic feature transformation $\phi(x_1) = x_1^2$.

c) Plot both regressions.

d) Which is a better fit a) or b)?

3 Thinking Questions

a) Until now we could only solve classification tasks where the two classes were separated by simple lines. Now we have seen that we can apply any feature transformations we want. Think about which kinds of problems we can solve now? Is it all a matter of finding the right transformation? Is it easy to choose the right transformation?

b) When using the linear regression for classification, think about how the threshold changes the sensitivity of the model. Is it more or less likely that the model will fail to recognize a class member as the threshold increases?