Lista Extra de Exercícios (2)
Introdução à Inteligência Artificial, CIC, UnB
Prof. Díbio

(Russel & Norvig, 2020) (cap. 19)

18.6 Consider the following data set comprised of three binary input attributes (A_1, A_2, and A_3) and one binary output:

Example	A_1	A_2	A_3	Output y
$\mathbf{x}_1$	1	0	0	0
$\mathbf{x}_2$	1	0	1	0
$\mathbf{x}_3$	0	1	0	0
$\mathbf{x}_4$	1	1	1	1
$\mathbf{x}_5$	1	1	0	1

Use the algorithm in Figure 18.5 (page 702) to learn a decision tree for these data. Show the computations made to determine the attribute to split at each node.

2. (Hastie et. al., 2021) (cap. 8)

5. Suppose we produce ten bootstrapped samples from a data set containing red and green classes. We then apply a classification tree to each bootstrapped sample and, for a specific value of X, produce 10 estimates of $P(\text{Class is Red}|X)$:

$$0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, \text{ and } 0.75.$$

There are two common ways to combine these results together into a single class prediction. One is the majority vote approach discussed in this chapter. The second approach is to classify based on the average probability. In this example, what is the final classification under each of these two approaches?

3. (Hastie et al., 2021) (cap. 8)

8. In the lab, a classification tree was applied to the `Carseats` data set after converting `Sales` into a qualitative response variable. Now we will seek to predict `Sales` using regression trees and related approaches, treating the response as a quantitative variable.

   (a) Split the data set into a training set and a test set.

   (b) Fit a regression tree to the training set. Plot the tree, and interpret the results. What test MSE do you obtain?

   (c) Use cross-validation in order to determine the optimal level of tree complexity. Does pruning the tree improve the test MSE?

   (d) Use the bagging approach in order to analyze this data. What test MSE do you obtain? Use the `importance()` function to determine which variables are most important.

   (e) Use random forests to analyze this data. What test MSE do you obtain? Use the `importance()` function to determine which variables are most important. Describe the effect of m, the number of variables considered at each split, on the error rate obtained.

4. (Mtichell, 1997) (cap. 3)

3.2. Consider the following set of training examples:

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	−	T	F
4	+	F	F
5	−	F	T
6	−	F	T

(*a*) What is the entropy of this collection of training examples with respect to the target function classification?

(*b*) What is the information gain of a_2 relative to these training examples?

5. Dada a seguinte tabela com dados de 8 pacientes, elabore baseando-se no algoritmo ID3 uma árvore de decisão. Mostre os passos e cálculos relacionados aos ganhos de informação quanto à escolha dos nós.

Paciente	**Febre**	**Dor Muscular**	**Manchas**	**Coceira**	**Diagnóstico**
P1	Alta	Frequente	Presente	Intensa	Emergência
P2	Moderada	Rara	Presente	Inexistente	Urgência
P3	Nenhuma	Frequente	Ausente	Inexistente	Mais_Exames
P4	Alta	Frequente	Ausente	Intensa	Emergência
P5	Moderada	Permanente	Presente	Inexistente	Urgência
P6	Moderada	Permanente	Ausente	Intensa	Urgência
P7	Alta	Permanente	Presente	Inexistente	Emergência
P8	Alta	Permanente	Ausente	Inexistente	Emergência
P9	Moderada	Frequente	Ausente	Moderada	Mais_Exames
P10	Moderada	Frequente	Ausente	Intensa	Emergência
P11	Nenhuma	Rara	Presente	Moderada	Mais_Exames
P12	Alta	Rara	Presente	Moderada	Urgência

What value of k minimizes the training set error for this dataset? What is ing training error?

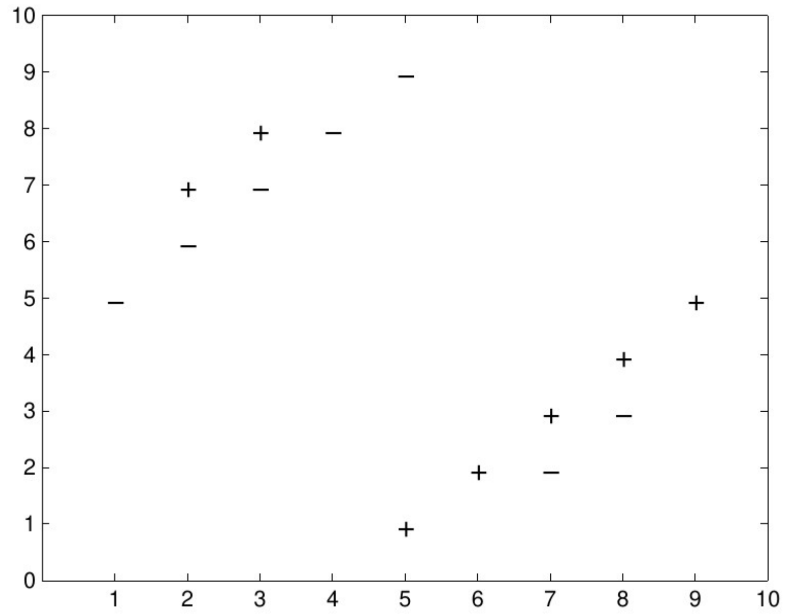


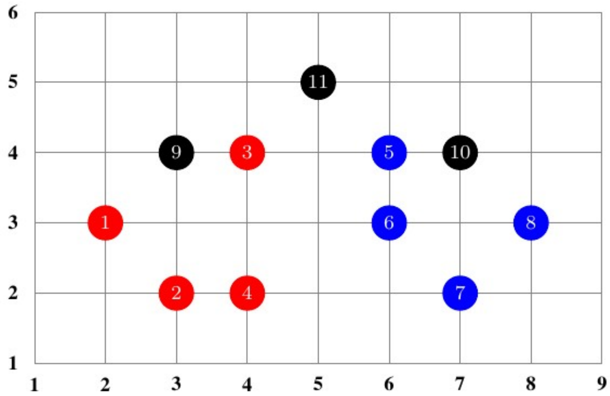
Figure 1: Dataset for KNN binary classification task.

6.

What value of k minimizes the training set error for this dataset? What is the resulting training error?

Why might using too large values k be bad in this dataset? Why might too small values of k also be bad?

7.



Consider the following training data:

$$x_1 = (2, 3), x_2 = (3, 2), x_3 = (4, 4), x_4 = (4, 2)$$

$$x_5 = (6, 4), x_6 = (6, 3), x_7 = (7, 2), x_8 = (8, 3)$$

Let $y_A = -1, y_B = +1$ be the class indicators for both classes

$$A = \{x_1, x_2, x_3, x_4\}, B = \{x_5, x_6, x_7, x_8\}.$$

(a) Just using the above-standing plot, specify which of the points should be identified as support vectors.

(b) Draw the maximum margin line which separates the classes (you don't have to do any computations here). Write down the normalized normal vector $\mathbf{w} \in \mathbb{R}^2$ of the separating line and the offset parameter $b \in \mathbb{R}$.

(c) Consider the decision rule: $H(x) = \langle \mathbf{w}, x \rangle + b$. Explain how this equation classifies points on either side of a line. Determine the class for the points $x_9 = (3, 4)$, $x_{10} = (7, 4)$ and $x_{11} = (5, 5)$.

8. (Russell & Norvig, 2020)

1.7 To what extent are the following computer systems instances of artificial intelligence:

- Supermarket bar code scanners.
- Web search engines.
- Voice-activated telephone menus.
- Internet routing algorithms that respond dynamically to the state of the network.

1.8 Many of the computational models of cognitive activities that have been proposed involve quite complex mathematical operations, such as convolving an image with a Gaussian or finding a minimum of the entropy function. Most humans (and certainly all animals) never learn this kind of mathematics at all, almost no one learns it before college, and almost no one can compute the convolution of a function with a Gaussian in their head. What sense does it make to say that the "vision system" is doing this kind of mathematics, whereas the actual person has no idea how to do it?

1.9 Why would evolution tend to result in systems that act rationally? What goals are such systems designed to achieve?

1.10 Is AI a science, or is it engineering? Or neither or both? Explain.

1.11 "Surely computers cannot be intelligent—they can do only what their programmers tell them." Is the latter statement true, and does it imply the former?

1.12 "Surely animals cannot be intelligent—they can do only what their genes tell them." Is the latter statement true, and does it imply the former?

1.13 "Surely animals, humans, and computers cannot be intelligent—they can do only what their constituent atoms are told to do by the laws of physics." Is the latter statement true, and does it imply the former?

15. Defina Aprendizado de Máquina.

16. Quais são os quatro (4) modelos de Aprendizado de Máquina?

17. Explique validação cruzada.

18. Calcule Acurácia, Precisão, Revocação e medida F1 da seguinte matriz de confusão de um classificador.

		Actual Condition		
		FALSE	TRUE	
Predicted Condition	FALSE	30	70	100
	TRUE	20	100	120
		50	170	220

Consider a classification problem in $\mathbb{R}^2$, with two classes, in which the training set is given by:

x_1	x_2	Class
–2	–2	A
-2	-1	A
1	2	A
2	1	A
-2	2	B
0	2	B
0	-1	B
2	-1	B

Consider the nonlinear mapping from input space to a two-dimensional feature space, given by

$$(x_1, x_2) \rightarrow (x_1^2, x_1 x_2)$$

a) Plot, on square-lined paper, the training patterns in input space, and label them according to the class they belong to. State whether the patterns from the two classes are linearly separable in this space.

b) Plot, on square-lined paper, separately from the graph made in step a), the training patterns in feature space, and label them according to the class they belong to.

c) Find the widest-margin classifier in feature space. More specifically, find the equations of the classification boundary and of the two margin boundaries. Plot these three boundaries on the same graph that was used in step b). Also indicate which are the support vectors in feature space. Note that the boundaries and the support vectors are easy to find by inspection.

d) Find which vectors, in input space, correspond to the support vectors found in step c).

e) Plot the classification boundary in input space, on the same graph that was used in step a).

Note that:

- Equations of the boundary in input space, in the forms $x_1 = f(x_2)$ and $x_2 = g(x_1)$, are easy to derive from the classification boundary equation found in step c).
- The boundary is a hyperbola whose asymptotes are easy to find from the first of those equations, by assuming that x_2 is large in absolute value.
- Using the second of those equations, it is easy to find points of the hyperbola for $x_1 = \pm 1/2$ and for small integer values of x_1 (both positive and negative).

f) Plot, on the same graph, the boundaries of the classification margin zone in input space. Shade the area between these two boundaries, to better visualize the classification margin zone in input space.

g) Using the graph completed in step f), check that the training patterns from different classes fall on different sides of the classification boundary, that the support vectors found in step d) fall on the margin boundaries, and that no training patterns fall within the classification margin zone.

h) Write the inequality that you would use to classify new input patterns with the SVM classifier developed in steps a) to g). Choose an inequality that is as simple as possible.

19.