Machine Learning Exercises 5

Due: November 22

Tim van Erven

November 15, 2007

Exercises

1. Compute the Euclidean distance between the following pairs of vectors:

   (a) $\begin{pmatrix} 3 \\ 1 \\ 8 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix}$

   (b) $\begin{pmatrix} 2 \\ -1 \\ 17 \\ 9 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ -1 \\ 17 \\ 9 \\ 1 \end{pmatrix}$

2. Consider a medical classification problem with two classes: Either a patient has a particular form of cancer or (s)he does not. Classifications are to be based on a single feature: the result of a medical test for this form of cancer. The test seems pretty accurate: Given that the patient really has cancer, it will indicate 'cancer' in 99% of the cases; Given that the patient does not have cancer, the test will indicate 'no cancer' in 98% of the cases. We will assume that 0.7% of the patients have cancer.[1]

   (a) If we use the notation from class, what are y and $\mathbf{x}$ in this example? What are their possible values? How many components does $\mathbf{x}$ have?

   (b) Let Ω be the set of possible values of $(y, \mathbf{x})^\top$ and let P be the distribution on Ω that corresponds to the introduction above. Let Y be a random variable that is 1 if a patient has cancer and 0 otherwise. Let X be a random variable that is 1 if the test result is 'cancer' and 0 otherwise. Please give the following probabilities: $P(X = 1 \mid Y = 1)$, $P(X = 1 \mid Y = 0)$ and $P(Y = 1)$.

   (c) Compute the probability that a patient has cancer given that the test indicates 'cancer'. Hint: You may use the fact that $P(X = 1) = P(X = 1 \wedge Y = 0) + P(X = 1 \wedge Y = 1)$.

   (d) Your answer to the previous question should imply that even when the test indicates cancer, it is still less likely that the patient really has cancer than that (s)he does not. How can this be? Which factor causes this to be the case?

[1] These numbers are entirely made up.

3. Suppose Ω is a sample space and that P is a distribution on Ω. Then the complement $\bar{A}$ of any event $A \subseteq \Omega$ is the event that A does **not** occur: $\bar{A} = \{\omega \mid \omega \in \Omega, \omega \notin A\}$. For any events A and B, it holds that $P(\bar{A}) = 1 - P(A)$ and $P(\bar{A} \mid B) = 1 - P(A \mid B)$.

   Suppose $A, B \subseteq \Omega$ are independent events under P. *Prove that A and $\bar{B}$ are independent events under P.*

   For simplicity, you may assume that $0 < P(A) < 1$ and $0 < P(B) < 1$ to avoid dividing by zero, although these assumptions are not strictly necessary.

Grading Policy

- Grades are between 1 and 10.
- You always start with 1 point.
- Partial points may be awarded for partially correct answers.