Answers Machine Learning Exercises 5

Tim van Erven

December 1, 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 $\begin{pmatrix} 2 \\ -1 \\ 17 \\ 9 \\ 1 \end{pmatrix}$

   Answers:

   (a) $\left\| \begin{pmatrix} 3 \\ 1 \\ 8 \end{pmatrix} - \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 5 \\ 0 \\ 5 \end{pmatrix} \right\| = \sqrt{5^2 + 0^2 + 5^2} = 5 \cdot \sqrt{2}$

   (b) $\left\| \begin{pmatrix} 2 \\ -1 \\ 17 \\ 9 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \\ 17 \\ 9 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\| = 0$

   Grading:

   - 1.5 points for each correct subquestion.
   - It is not possible to get partial points on this question.

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?

[1] These numbers are entirely made up.

(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?

Answers:

(a) y is the class label: Does the patient have cancer or not. $\mathbf{x}$ is a 1-dimensional feature vector. Its single component x_1 indicates whether the test result for a patient is 'cancer' or 'no cancer'.

(b)

$$P(X = 1 \mid Y = 1) = 0.99 \quad P(X = 1 \mid Y = 0) = 0.02 \quad P(Y = 1) = 0.007$$

(c) We have to apply Bayes' rule:

$$\begin{aligned}
P(Y = 1 \mid X = 1) &= \frac{P(X = 1 \mid Y = 1)P(Y = 1)}{P(X = 1)} \\
&= \frac{P(X = 1 \mid Y = 1)P(Y = 1)}{P(X = 1 \wedge Y = 0) + P(X = 1 \wedge Y = 1)} \\
&= \frac{P(X = 1 \mid Y = 1)P(Y = 1)}{P(X = 1 \mid Y = 0)P(Y = 0) + P(X = 1 \mid Y = 1)P(Y = 1)} \\
&= \frac{0.99 \cdot 0.007}{0.02 \cdot 0.993 + 0.99 \cdot 0.007} \\
&\approx 0.259
\end{aligned}$$

(d) The reason is that such a small percentage (0.7%) of the whole patient population has cancer. So before hearing the result of the test, we have a pretty strong belief that a patient does not have cancer. Notice that conditioning on the test result 'cancer' does increase the probability of really having cancer significantly. So after hearing the test result, we will certainly want to run more tests.

Grading:

- 1 point for each correct subquestion.
- Partial points may be awarded.
- If you make a mistake in subquestion (b), then subquestion (c) is (only) considered correct if it consistently uses your earlier wrong result.

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.

   Answer: We have to prove that

$$P(A \cap \bar{B}) = P(A)P(\bar{B}).$$

To do this, we will have to rewrite $P(A \cap \bar{B})$ into an expression containing $P(A \cap B)$, because then we can apply our assumption that A and B are independent. That means that we will first have to rewrite probabilities involving $\bar{B}$ into probabilities involving B. We can do this as follows:

$$\begin{aligned}
P(A \cap \bar{B}) &= P(\bar{B} \mid A) \cdot P(A) \\
&= (1 - P(B \mid A)) \cdot P(A) \\
&= P(A) - P(B \mid A) \cdot P(A) \\
&= P(A) - P(A \cap B) \\
&= P(A) - P(A)P(B) \\
&= P(A) \cdot (1 - P(B)) \\
&= P(A)P(\bar{B}).
\end{aligned}$$

   Grading:

   - 2 points for any correct answer.
   - If the answer is not fully correct, then you get
     - 0.5 points for rewriting $P(A \cap \bar{B})$ or $P(\bar{B})$ into an expression involving B instead of $\bar{B}$.
     - 1 point for using the independence of A and B in a useful way, i.e. applying $P(A \cap B) = P(A)P(B)$.

     If you did not get the partial points above, you may also get partial points for other clearly identifiable progress you have made in solving the question.

Grading Policy

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