Machine Learning Exercises 3

Due: October 18

Tim van Erven

October 12, 2007

Exercises

1. In many fairy tales it is possible to determine whether a certain character is good or evil based on the features in Table 1. This is a classification task. Let $\mathbf{x}$ be a 3-dimensional feature vector that contains the features from Table 1: x_1 corresponds to WearsBlack, x_2 to SavesPrincess, and x_3 to HorseColour. Let y be the label 'Good' or 'Evil'.

Table 1: Fairy tale features

Feature	Possible Values	Description
WearsBlack	Yes, No	Does the character wear black clothes?
SavesPrincess	Yes, No	Does the character save a princess?
HorseColour	Black, White, Brown	What is the colour of the character's horse?

Let Y be the random variable that is defined as follows:

$$Y\left(\begin{pmatrix} y \\ \mathbf{x} \end{pmatrix}\right) = \begin{cases} 1 & \text{if } y = \text{Evil}, \\ 2 & \text{if } y = \text{Good}. \end{cases}$$

(a) *Write down the definition of the entropy of Y. (Not the one from Mitchell, but the one from class. In Mitchell's version probabilities have already been replaced by their estimates.)*

The definition of $H(Y)$ contains some probabilities.

(b) *Estimate these probabilities from the data in Table 2 and use your estimates to compute $H(Y)$.*

Table 2: Fairy tale data set

x_1 WearsBlack	x_2 SavesPrincess	x_3 HorseColour	y GoodOrEvil
No	Yes	Black	Good
Yes	No	Black	Evil
No	No	White	Good
Yes	Yes	Brown	Good

Let X_3 be the random variable that is defined as follows:

$$X_3\left(\begin{pmatrix} y \\ \mathbf{x} \end{pmatrix}\right) = \begin{cases} 1 & \text{if } x_3 = \text{Black}, \\ 2 & \text{if } x_3 = \text{White}, \\ 3 & \text{if } x_3 = \text{Brown}. \end{cases}$$

(c) *Write down the definition of* $H(Y \mid X_3 = 1)$.

The conditional probability of, for example, $P(Y = 1 \mid X_3 = 1)$ may be estimated indirectly: We estimate $P(Y = 1 \text{ and } X_3 = 1)$ and $P(X_3 = 1)$, and then use the definition of conditional probability as follows:

$$P(Y = 1 \mid X_3 = 1) = \frac{P(Y = 1 \text{ and } X_3 = 1)}{P(X_3 = 1)}.$$

Other conditional probabilities can be estimated in the same way.

(d) *Compute* $H(Y \mid X_3 = 1)$, $H(Y \mid X_3 = 2)$ *and* $H(Y \mid X_3 = 3)$ *using estimates of the relevant conditional probabilities from Table 2.*

(e) *Write down the definition of the conditional entropy of* Y *given* X_3 *and compute it using probability estimates from Table 2.*

(f) *Write down the definition of the mutual information* $I(Y; X_3)$ *between* Y *and* X_3 *and compute it.*

The estimate of the mutual information between Y and X_3 that you have just computed is called the information gain of attribute X_3 by Mitchell.

2. Consider the dice prediction game that we played in class (see slide 16 from `mlslides6.pdf`). Suppose we played this game with fewer students. *Would the risk of overfitting our train set go up, down or stay the same?*

Grading Policy

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