Chalmers tekniska högskola

TDA231 *Algorithms for Machine Learning & Interference*
Teacher: *Devdatt Dubhashi*

Homework 0

Victor Nilsson - *vicnilss@student.chalmers.se*

17 januari 2017

1 Theoretical problems

Problem 1.1

The probability that a random person are cancer positive is $P(C_+) = 0.0001$ and thus the probability of being cancer negative is $P(C_-) = 0.9999$. The testing of the disease is not 100 % accurate and the probabiliy of actually getting a positive test given the patient being sick is $P(T_+|C_+) = 0.99$ and conversely the probability of getting negative result when being healthy is $P(T_-|C_-) = 0.99$. The definition of conditional probability is,

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}, \tag{1}$$

so the risk of actually being sick given a positive result is,

$$\begin{array}{rcl} P(C_+|T_+) & = & \frac{P(C_+ \cap T_+)}{P(T_+)} \\ & \Rightarrow & \\ P(C_+ \cap T_+) & = & P(C_+|T_+)P(T_+), \end{array} \tag{2}$$

and similarly we can write,

$$P(C_+ \cap T_+) = P(T_+|C_+)P(C_+), \tag{3}$$

and thus,

$$\begin{array}{rcl} P(T_+|C_+)P(C_+) & = & P(C_+|T_+)P(T_+) \\ & \Rightarrow & \\ P(C_+|T_+) & = & \frac{P(T_+|C_+)P(C_+)}{P(T_+)}. \end{array} \tag{4}$$

We can also use that

$$\begin{array}{rcl} P(T_+) & = & P(T_+|C_+)P(C_+) + P(T_+|C_-)P(C_-) \\ & = & P(T_+|C_+)P(C_+) + (1 - P(T_-|C_-))P(C_-) \end{array} \tag{5}$$

Inserting the given data we get,

$$\begin{array}{rcl} P(C_+|T_+) & = & \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.01 \cdot 0.9999} \\ & \approx & 0.0098 \end{array} \tag{6}$$

Answer: So the chance of actually being sick is only a mere 0.98 %.

Problem 1.2

The covariance of two random variables X and Y is defined as,

$$\begin{array}{rcl} Cov[X,Y] & \triangleq & E[(X - E[X])(Y - E[Y])] \\ & = & E[XY] - E[X]E[Y] \end{array} \tag{7}$$

.

For our problem we have that X is uniform on the interval $[-1,1]$ and $Y := X^2$, so

$$
\begin{array}{rcl}
E[X] & = & \int_{-1}^{1} x p_X(x) dx \\
\left\{ \begin{array}{rcl} \int_{-1}^{1} p(x)_X dx & = & 1 \\ \int_{-1}^{1} c dx & = & 1 \\ 2c & = & 1 \\ c & = & \frac{1}{2} \end{array} \right\} & = & \frac{1}{2} \int_{-1}^{1} x dx \\
& = & \frac{1}{4} \left[x^2 \right]_{-1}^{1} \\
& = & 0.
\end{array} \tag{8}
$$

Since $E[X]$ is zero we do not have to calculate the expectation of Y since it disappears anyway, though, we need to calculate the expectation of the product,

$$
\begin{array}{rcl}
E[XY] & = & \int_{-1}^{1} x p_{XY}(x) dx \\
\left\{ \; p_{XY}(x) \;\; \propto \;\; p_X^3(x) \; \right\} & \propto & \int_{-1}^{1} x p_X^3(x) dx \\
& = & 0.
\end{array} \tag{9}
$$

If we insert this in Eq. (7) we get

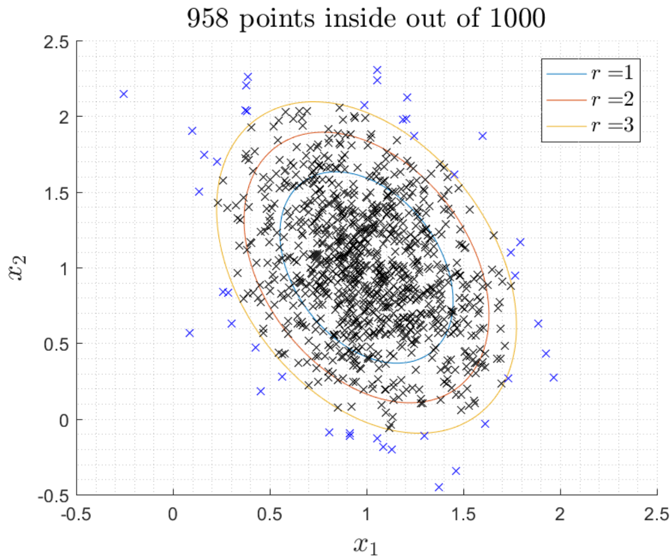
$$
\begin{array}{rcl}
Cov[X,Y] & = & 0 - 0 \cdot E[Y] \\
& = & 0
\end{array} \tag{10}
$$

This means that the random variables X and Y are *uncorrelated* which does not imply *independence* since the variables obviously are.

2 Practical problems

Problem 2.1

$$ \tag{11} $$



Figur 1: *Level set for the function given in Eq.* (11).

Problem 2.2