Machine Learning
Name: Duong Thi Thu Phuong

HOMEWORK 1

Question 1. *To evaluate a new test for detecting Hansen's disease, a group of people 5% of which are known to have Hansen's disease are tested. The test finds Hansen's disease among 98% of those with the disease and 3% of those who don't. What is the probability that someone testing positive for Hansen's disease under this new test actually has it?*

My Answer 1. *Suppose:*
X : are known to have Hansen's disease.
Y : are known not to have Hansen's disease.
D : result of the tested is positive
So:
P(X) = 0.05 ⇒ P(Y) = 1- 0.05 = 0.95
P(D|X) = 0.98
P(D|Y) = 0.03
The probability that someone testing positive for Hansen's disease under this new test actually is:

$$P(X|D) = \frac{P(D|X).P(X)}{P(D)} = \frac{P(D|X).P(X)}{P(D|X).P(X) + P(D|Y).P(Y)} = \frac{0.98 \cdot 0.05}{0.98 \cdot 0.05 + 0.03 \cdot 0.95} \cong 63.22\%$$

Question 2. *Proof the following distributions are normalized then calculate the mean and standard deviation of these distribution:*
a, Univariate normal distribution.
b, (Optional) Multivariate normal distribution.

My Answer 2. *a, We have univariate normal distribution:*

$$P(X) = \frac{1}{\sqrt{2\pi\sigma^2}}.e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Set: $a = x - \mu \Rightarrow da = dx$
Univariate normal distribution are normalized when:

$$\leftrightarrow \int P(x)dx = 1$$

So:

$$N = \int \frac{1}{\sqrt{2\pi\sigma^2}}.e^{\frac{-a^2}{2\sigma^2}} da = 1$$

$$\leftrightarrow N = \int e^{\frac{-a^2}{2\sigma^2}} da = \sqrt{2\pi\sigma^2}$$

$$\leftrightarrow N = \int \int e^{\frac{-a^2}{2\sigma^2}} e^{\frac{-b^2}{2\sigma^2}} dadb = 2\pi\sigma^2$$

Set: $a = xcos\theta$ *and* $b = xsin\theta$
Then:

$$dadb = \begin{vmatrix} \frac{\partial a}{\partial x} & \frac{\partial a}{\partial \theta} \\ \frac{\partial b}{\partial x} & \frac{\partial a}{\partial \theta} \end{vmatrix} dxd\theta = \begin{vmatrix} cos\theta & -xsin\theta \\ sin\theta & xcos\theta \end{vmatrix} dxd\theta = (xcos^2\theta + xsin^2\theta)dxd\theta$$

$$\Rightarrow dadb = xdxd\theta$$

$$N = \int_0^{2\pi} \int_0^{+\infty} e^{\frac{-x^2}{2\sigma^2}} xdxd\theta$$

$$= \int_0^{2\pi} -\sigma^2 e^{\frac{-x^2}{2\sigma^2}} |_0^{+\infty} d\theta$$

$$= \int_0^{2\pi} \sigma^2 d\theta \quad (Because \quad e^{+\infty} = 0) = 2\pi\sigma^2$$

$$\Rightarrow Univariate \quad normal \quad distribution \quad normalization$$