Machine Learning 1
Homework 1

Alivanistos Dimitris

September 14, 2016

1 Question 1

1.1 Rain

1. Define the random variables and the values they can take on, both with symbols and numerically.
   Proof.

   Considering the above we have 3 Random Variables.

   A: The possibility of being in Amsterdam with values $A = \{'True', 'False'\}$ or $A = \{0, 1\}$

   RT: The possibility of being in Rotterdam with values $RT = \{'True', 'False'\}$ or $RT = \{0, 1\}$

   R: The possibility that it's raining with values $R = \{'True', 'False'\}$ or $R = \{0, 1\}$

   Given the above we denote the following conditional probabilities and priors:

   $$P(A = 1) = 0.8$$
   $$P(RT = 1) = 0.2$$
   $$P(R = 1 \mid A = 1) = 0.5$$
   $$P(R = 1 \mid RT = 1) = 0.75$$

   ■

2. What is the probability that it does not rain when you are in Rotterdam?

   Proof.
   The probability of not raining when in Rotterdam is the complement of raining when in Rotterdam. So the required probability can be calculated by subtracting the complement which is given from 1 as shown below:
   $P(R = 0 \mid RT = 1) = 1 - P(R = 1 \mid RT = 1) = 1 - 0.75 = 0.25$ ■

3. What is the probability that it rains where you are?
   Proof.

   We need to calculate the total probability of P(R):

   $$\begin{aligned} P(R) &= P(R = 1 \mid A = 1)P(A = 1) + P(R = 1 \mid RT = 1)P(RT = 1) \\ &= 0.5 \cdot 0.8 + 0.2 \cdot 0.75 = 0.55 \end{aligned}$$

   ■

4. You wake up on the sidewalk, after a night out which you can't remember anything about but which clearly was not such a great idea. You can't recognize your surroundings, but you must be either in Amsterdam or Rotterdam. It is raining. What is the probability that you are in Amsterdam?
   Or P(A=1 | R=1)?
   Proof.

   We will use the Baye's Theorem in order to calculate the desired probability:

   $$P(A = 1 \mid R = 1) = \frac{P(R = 1 \mid A = 1)P(A = 1)}{P(R = 1)} = \frac{0.5 \cdot 0.8}{0.55} = 0.727$$

   ■

1.2 Cancer

1. What is p(cancer) and p(not cancer)?
   Variables :
   C: The probability of having cancer with values $C = \{'True', 'False'\}$ or $C = \{0, 1\}$

   T: The probability of the test being positive or negative with values $T = \{'Positive', 'Negative'\}$ or $T = \{0, 1\}$

   P: The total population of the city with value $P = 500000$.

   PC: The population that has cancer with value $PC = 500$.

   Derived prior and conditional probabilities

   $$\begin{aligned} &P(C = 1) = \frac{PC}{P} = \frac{500}{500000} = 0.001 \\ &P(C = 0) = 1 - P(C = 1) = 0.999 \\ &P(T = 1 \mid C = 1) = 0.99 \\ &P(T = 1 \mid C = 0) = 0.05 \end{aligned}$$

2. If a person takes the blood test and it returns positive, what is the probability the patient has cancer? Calculate marginal probability of $p(T=1)$

$$
\begin{aligned}
&p(T=1)\\
&\qquad = p(T=1|C=1) * p(C=1)\\
&\qquad + p(T=1|C=0) * p(C=0)\\
&\qquad = 0.99 * 0.001 + 0.05 * 0.999 = 0.05094
\end{aligned}
$$

Using Bayes Rule:

$$p(C=1|T=1) = \frac{p(T=1|C=1) * p(C=1)}{p(T=1)} = \frac{0.99 * 0.001}{.5094} = 0.1943$$

3. What are some of the assumptions we are implicitly making when answering this question?

   A basic assumption is that everyone in the population will take the test. Another assumption is that the people who do have cancer are 1 - P(C = 0) = 0.001 which means that 1 every 1000 people have cancer (with an insufficient sample of data). Last but not least we are assuming that people who have cancer and people who do not, show the same symptoms, which usually in real life is not the case.

1.3 Posterior Normal distribution

1. Write down the general expression for a posterior distribution, using for the parameter θ, D for the data. Indicate the prior, likelihood, evidence, and posterior.
   According to the Bayes Theorem a posterior distribution is written as shown below. Using θ and D we get:

$$p(\theta|D) = \frac{p(D|\theta) * p(\theta)}{p(D)} \tag{1}$$

   where $p(\theta|D)$ is the **posterior**
   $p(D|\theta)$ is the **likelihood**
   $p(\theta)$ is the **prior**
   $p(D)$ is the **evidence**
   When $p(D)$ is not known we can actually rewrite it as a marginalization of a continuous variable. Therefore:

$$p(\theta|D) = \frac{p(D|\theta) * p(\theta)}{\int_{\theta_0=\theta} p(D|\theta_0) * p(\theta_0)} \tag{2}$$

2. Write the posterior for this particular example. You do not need an analytic solution.
   Taking the previous equation 2, and substituting the particular Normal distributions given in this example, we have:
   $p(\theta|D) = p(\mu|D)$
   $p(D|\theta) = p(D|\mu) = \prod_{n=1}^{N} N(x_n|\mu, \sigma^2)$
   $p(\theta) = p(\mu) = N(\mu|\mu_0, \sigma_0^2)$
   $p(D) = \int_\mu N(\mu|\mu_0, \sigma_0^2) * \prod_{n=1}^{N} N(x_n|\mu, \sigma^2)$

$$p(\theta|D) = \frac{p(D|\theta) * p(\theta)}{\int_{\theta_0=\theta} p(D|\theta_0) * p(\theta_0)}$$
$$= \frac{N(\mu|\mu_0, \sigma_0^2) * \prod_{n=1}^{N} N(x_n|\mu, \sigma^2}{\int_{\mu} N(\mu|\mu_0, \sigma_0^2) * \prod_{n=1}^{N} N(x_n|\mu, \sigma^2)}$$

Also, we know that the posterior will be a Gaussian given that the prior is a Gaussian as well.

2 Basic Linear Algebra and Derivatives

2.1 Matrices

Let A $= \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix}$ and b $= \begin{bmatrix} 9 \\ 5 \end{bmatrix}$

1. Compute **Ab**

   Proof.

   $\boldsymbol{Ab} = \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \cdot 9 + 5 \cdot 5 \\ 2 \cdot 9 + 3 \cdot 5 \end{bmatrix} = \begin{bmatrix} 27 + 25 \\ 18 + 15 \end{bmatrix} = \begin{bmatrix} 52 \\ 33 \end{bmatrix}$ ■

2. Compute $\boldsymbol{b^T A}$

   Proof.

   $\boldsymbol{b^T A} = \begin{bmatrix} 9 & 5 \end{bmatrix} \cdot \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 9 * 3 + 5 * 2 & 9 * 5 + 5 * 3 \end{bmatrix} = \begin{bmatrix} 37 & 60 \end{bmatrix}$ ■

3. What is the vector **c** for which $\boldsymbol{Ac = b}$

   To solve the linear system we have the following equation:
   $Ac = b \Rightarrow \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} c_1 & c_2 \end{bmatrix} = \begin{bmatrix} 9 & 5 \end{bmatrix} \Rightarrow \begin{cases} 3 \cdot c_1 + 5 \cdot c_2 = 9 \\ 2 \cdot c_1 + 3 \cdot c_2 = 5 \end{cases}$
   Solving w.r.t. c_1 on the top equation:

$$3 \cdot c_1 + 5 \cdot c_2 = 9 \Rightarrow c_1 = 3 - \frac{5}{3} c_2 \tag{3}$$

   Replacing c_1 from (1) in the 2nd equation results in:

$$2 \cdot (3 - \frac{5}{3} \cdot c_2) + 5 \cdot c_2 = 5$$
$$\Rightarrow 6 - \frac{10}{3} \cdot c_2 + 3 \cdot c_2 = 5$$
$$\Rightarrow 18 - 10 \cdot c_2 + 9 \cdot c_2 = 15 \Rightarrow \cdot c_2 = -3$$

   Using $\boldsymbol{c_2 = 3}$ on (1) we get that $\boldsymbol{c_1 = -2}$
   So $\boldsymbol{c} = \begin{bmatrix} -2 & 3 \end{bmatrix}$

4. What is $\boldsymbol{A^{-1}}$

   For a 2x2 matrix the Inverse is: $\boldsymbol{A^{-1}} = \frac{1}{\mathbf{det(A)}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$ where $\det(A) =| a_{11} \cdot a_{22} - a_{12} \cdot a_{21} |$which is the determinant.

   So we can calculate the inverse of **A** as: $\boldsymbol{A^{-1}} = \frac{1}{3 \cdot 3 - 5 \cdot 2} \begin{bmatrix} 3 & -5 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -3 & 5 \\ 2 & -3 \end{bmatrix}$

5. Verify that $\boldsymbol{A^{-1}b = c}$. Show that this must be the case.

   $\begin{bmatrix} -3 & -5 \\ 2 & -3 \end{bmatrix} * \begin{bmatrix} 9 \\ 5 \end{bmatrix} \mathbf{=} \begin{bmatrix} -3 * 9 + 5 * 5 \\ 2 * 9 + 5 * (-3) \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \mathbf{= c}$

2.2 Derivatives

1. $x^2 + 2x + 3$
   Proof.
   Let $f(x) = x^2 + 2x + 3$. The gradient is:
   $f^{'}(x) = (x^2 + 2x + 3)' = 2x + 2$ ■

2. $(2x^3 + 1)^2$
   Proof.
   Let $f(x) = (2x^3 + 1)^2$. The gradient is:

   $$\begin{aligned} f^{'}(x) &= (2x^3 + 1)^{2^{'}} = 2(2x^3 + 1)(2x^3 + 1)^{'} \\ &= (4x^3 + 2)(3 \cdot 2x^2) \\ &= 4x^3 6x^2 + 2 \cdot 6x^2 = 24x^5 + 12x^2 \end{aligned}$$

   ■

3. $f(x, y, z) = (x + 2y)^2 sin(xy)$

   Proof.

   First we will find the partial derivative with respect to x.

   $$\begin{aligned} \frac{\partial f}{\partial x} &= 2(x + 2y)\frac{\partial(x + 2y)}{\partial x} * \sin(xy) + (x + 2y)^2 \frac{\partial \sin(xy)}{\partial x} && (4) \\ &= 2(x + 2y)\sin(xy) + y(x + 2y)^2 \cos(xy) && (5) \end{aligned}$$

   Then we will find the partial derivative with respect to y.

   $$\begin{aligned} \frac{\partial f}{\partial y} &= 2(x + 2y)\frac{\partial(x + 2y)}{\partial y} * \sin(xy) + (x + 2y)^2 \frac{\partial \sin(xy)}{\partial x} && (6) \\ &= 4(x + 2y)\sin(xy) + x(x + 2y)^2 \cos(xy) && (7) \end{aligned}$$

   Last we will find the partial derivative with respect to z which is:

   $$\frac{\partial f}{\partial z} = 0 \tag{8}$$

   ■

4. $f(x, y, z) = 2\log(x + y^2 - z)$

   Proof.

   First we will find the partial derivative with respect to x.

$$\frac{\partial f}{\partial x} = \frac{2}{x + y^2 - z} \frac{\partial(x + y^2 - z)}{\partial x} = \frac{2}{x + y^2 - z} \tag{9}$$

   Then we will find the partial derivative with respect to y.

$$\frac{\partial f}{\partial y} = \frac{2}{x + y^2 - z} \frac{\partial(x + y^2 - z)}{\partial y} = \frac{2 * 2y}{x + y^2 - z} = \frac{4y}{x + y^2 - z} \tag{10}$$

   Last we will find the partial derivative with respect to z which is:

$$\frac{\partial f}{\partial z} = \frac{2 * -1}{x + y^2 - z} \tag{11}$$

   ■

5. $f(x, y, z) = \exp(x\cos(y + z))$

   Proof.

   First we will find the partial derivative with respect to x.

$$\frac{\partial f}{\partial x} = \exp(x\cos(y + z)) \frac{\partial(x\cos(y + z))}{\partial x} = \exp(x\cos(y + z))\cos(y + z) \tag{12}$$

   Then we will find the partial derivative with respect to y.

$$\frac{\partial f}{\partial y} = \exp(x\cos(y + z)) \frac{\partial(x\cos(y + z))}{\partial y} = \exp(x\cos(y + z))[x(\sin(y + z)) \frac{\partial(y + z)}{\partial y}] \tag{13}$$

$$= \exp(x\cos(y + z))x * \sin(y + z) * 1 \tag{14}$$

   Last we will find the partial derivative with respect to z which is:

$$\frac{\partial f}{\partial z} = \exp(x\cos(y + z)) \frac{\partial(x\cos(y + z))}{\partial z} = \exp(x\cos(y + z))[-x(\sin(y + z)) \frac{\partial(y + z)}{\partial z}] \tag{15}$$

$$= \exp(x\cos(y + z))(-x * \sin(y + z)) \tag{16}$$

   ■

2.3 Vectors Matrices

Given the following expression:

$(x-\mu)^T\Sigma^{-1}(x-\mu)+(\mu-\mu_0)^T S^{-1}(\mu-\mu_0)$

Answer the following:

1. Expand the expression and gather terms:

   Proof.
   $(x-\mu)^T\Sigma^{-1}(x-\mu)+(\mu-\mu_0)^T S^{-1}(\mu-\mu_0))$
   $= x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu$
   $+ \mu^T S^{-1}\mu - \mu^T S^{-1}\mu_0 - \mu_0^T S^{-1}\mu + \mu_0^T S^{-1}\mu_0$
   ■

   With S^{-1}, Σ^{-1} being symmetric and invertible matrices we can use the following properties:

   $\mu^T\Sigma^{-1}x = x^T\Sigma^{-1}\mu$ and $\mu_0^T S^{-1}\mu = \mu^T S^{-1}\mu_0$

   Given these 2 properties we can rewrite the above as:
   Proof.
   $x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu + \mu^T S^{-1}\mu - \mu^T S^{-1}\mu_0 - \mu_0^T S^{-1}\mu + \mu_0^T S^{-1}\mu_0$
   $= x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu - 2\mu^T\Sigma^{-1}x + \mu^T S^{-1}\mu + \mu_0^T S^{-1}x\mu_0 - 2\mu^T S^{-1}\mu_0$ ■

2. Collect all the terms that depend on μ and those that do not.
   Proof.
   $x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}\mu - 2\mu^T\Sigma^{-1}x + \mu^T S^{-1}\mu + \mu_0^T S^{-1}x\mu_0 - 2\mu^T S^{-1}\mu_0$
   $= \mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x + S^{-1}\mu_0) + x^T\Sigma^{-1}x + \mu_0^T S^{-1}\mu_0$

   Clearly our first 2 terms are the ones that depend on μ and the other 2 are not. ■

3. Take the derivative with respect to μ set to 0 , and solve for μ
   Proof.

   $\frac{\partial(\mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x+S^{-1}\mu_0)+x^T\Sigma^{-1}x+\mu_0^T S^{-1}\mu_0)}{\partial\mu} = 0$
   Since we take the derivative with respect to μ the terms that do not depend on μ are equal to 0.
   So:

   $\frac{\partial(\mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x+S^{-1}\mu_0)}{\partial\mu} = 0$

   $(\Sigma^{-1}+S^{-1})\mu + (\mu^T(\Sigma^{-1}+S^{-1}))^T - 2(\Sigma^{-1}x + S^{-1}\mu_0) = 0$

   $(\Sigma^{-1}+S^{-1})\mu + (\Sigma^{-1}+S^{-1})\mu - 2(\Sigma^{-1}x + S^{-1}\mu_0) = 0$

$$2(\Sigma^{-1} + S^{-1})\mu - 2(\Sigma^{-1}x + S^{-1}\mu_0) = 0$$

$$2(\Sigma^{-1} + S^{-1})\mu = 2(\Sigma^{-1}x + S^{-1}\mu_0)$$

$$(\Sigma^{-1} + S^{-1})\mu = (\Sigma^{-1}x + S^{-1}\mu_0)$$

$$\mu = (\Sigma^{-1} + S^{-1})^{-1}(\Sigma^{-1}x + S^{-1}\mu_0)$$

■