Assignment 1
Machine Learning 1, Fall 2016

Dana Kianfar
University of Amsterdam

September 14, 2016

1 Probability Theory

Question 1.1

1. We identify two discrete random variables W for the weather and L for the location defined as follows.

$$W \in \{r = rain, nr = no\ rain\}, L \in \{a = Amsterdam, r = Rotterdam\}$$

Which can take the following numerical values.

$$P(W = r \mid L = a) = 0.5$$
$$P(W = r \mid L = r) = 0.75$$
$$P(L = a) = 0.8$$
$$P(L = r) = 0.2$$

2.
$$P(W = nr \mid L = r) = 1 - P(W = r \mid L = r) = 0.25$$

3.
$$P(W = r) = \sum_l P(W = r \mid L = l)P(L = l)$$
$$P(W = r \mid L = a) \cdot P(L = a) + P(W = r \mid L = r) \cdot P(L = r)$$
$$0.5 \times 0.8 + 0.75 \times 0.2 = 0.55$$

4. Using the Bayes rule, we find the following.

$$P(L = a \mid W = r) = \frac{P(W = r \mid L = a) \cdot P(L = a)}{P(W = r)}$$
$$\frac{0.5 \times 0.8}{0.55} = 0.\overline{72}$$

Question 1.2

We formulate the problem using two discrete random variables C for cancer and D for diagnosis. They are defined as follows.

$$C \in \{c = cancer, nc = not\ cancer\}$$
$$D \in \{p = positive, n = negative\}$$

1.

$$P(C = c) = \frac{500}{500\,000} = 0.001$$

$$P(C = nc) = 1 - P(C = c) = 1 - 0.001 = 0.999$$

2. We are given the following information regarding the accuracy of the blood test.

$$P(D = p \mid C = c) = 0.99 \qquad \text{true positives}$$

$$P(D = n \mid C = c) = 0.01 \qquad \text{false negatives}$$

$$P(D = p \mid C = nc) = 0.05 \qquad \text{false positives}$$

$$P(D = n \mid C = nc) = 0.95 \qquad \text{true negatives}$$

We can then calculate the solution as follows.

$$P(C = c \mid D = p) = \frac{P(D = p \mid C = c) \cdot P(C = c)}{P(D = p)}$$

$$P(C = c \mid D = p) = \frac{P(D = p \mid C = c) \cdot P(C = c)}{\sum_x P(D = p \mid C = x) \cdot P(C = x)}$$

$$P(C = c \mid D = p) = \frac{P(D = p \mid C = c) \cdot P(C = c)}{P(D = p|C = c) \cdot P(C = c) + P(D = p|C = nc) \cdot P(C = nc)}$$

$$\frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times 0.999} = 0.01943$$

3. We assume that the people taking the blood test are sampled in the same way from the overall population as the people who are considered to have cancer. To elaborate, we assume that the events of a person taking the blood test and having cancer are independent. We also assume that both samples are independently and identically distributed within the population. Any prior knowledge that a person may be prone to cancer makes them more likely to take the blood test, and if the initial estimate of 500 cancer patients was based on data collected at in an unrepresentive or special setting, say in proximity of a nuclear facility or chemical factory in the city, then our measurements are likely biased.

Question 1.3

1. The posterior is calculated as follows.

$$\textit{posterior} = \frac{\textit{likelihood} \times \textit{prior}}{\textit{evidence}}$$

We replace each term by its mathematical notation as follows.

$$P(\theta \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \theta) \cdot P(\theta)}{P(\mathcal{D})}$$

Where:

$P(\theta \mid \mathcal{D})$: is the posterior

$P(\mathcal{D} \mid \theta)$: is the likelihood

$P(\theta)$: is the prior

$P(\mathcal{D})$: is the evidence

2. Similarly, for the specific problem at hand, we have the following.

$$P(\mu \mid \mathcal{D}) = \frac{\prod_i \mathcal{N}(x_i \mid \mu, \sigma^2) \cdot \mathcal{N}(\mu \mid \mu_0, \sigma_0^2)}{\int \left(\prod_i \mathcal{N}(x_i \mid \mu, \sigma^2)\right) \cdot \mathcal{N}(\mu \mid \mu_0, \sigma_0^2) d\mu}$$

Where:

$P(\mu \mid \mathcal{D})$: is the posterior
$P(\mathcal{D} \mid \mu) = \prod_i P(x_i \mid \mu) = \prod_i \mathcal{N}(x_i \mid \mu, \sigma^2)$: is the likelihood
$P(\mu) = \mathcal{N}(\mu \mid \mu_0, \sigma_0^2)$: is the prior
$P(\mathcal{D}) = \int P(\mathcal{D} \mid \mu) \cdot P(\mu) d\mu = \int \prod_i \mathcal{N}(x_i \mid \mu, \sigma^2) \cdot \mathcal{N}(\mu \mid \mu_0, \sigma_0^2) d\mu$: is the evidence

2 Basic Linear Algebra and Derivatives

Question 2.1

1.

$$Ab = \begin{bmatrix} 3 \times 9 + 5 \times 5 \\ 2 \times 9 + 3 \times 5 \end{bmatrix} = \begin{bmatrix} 52 \\ 33 \end{bmatrix}$$

2.

$$b^T = \begin{bmatrix} 9 & 5 \end{bmatrix}$$

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

3.

$$Ac = b \qquad c = \begin{bmatrix} x \\ y \end{bmatrix}$$

We can find a solution by solving it as a linear system as follows.

$$\begin{cases} 3x + 5y = 9 \\ 2x + 3y = 5 \end{cases} \xrightarrow[\times -3]{\times 2} \begin{cases} 6x + 10y = 18 \\ -6x - 9y = -15 \end{cases}$$

$$\begin{array}{cccc} & 6x & +10y= & 18 \\ + & -6x & -9y= & -15 \\ \hline & 0x & +y= & 3 \end{array}$$

It follows that $x = -2$, and therefore $c = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$

4.

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \det X = ad - bc$$

If $\det X \neq 0$ then an inverse exists as follows: $X^{-1} = \frac{1}{|\det X|} \cdot \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

For matrix A we can therefore calculate the following.

$$\det A = 3 \times 3 - 5 \times 2 = -1 \neq 0$$

$$A^{-1} = -1 \cdot \begin{bmatrix} 3 & -5 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -3 & 5 \\ 2 & -3 \end{bmatrix}$$

5.

$$A^{-1}b = \begin{bmatrix} -3 & 5 \\ 2 & -3 \end{bmatrix} \begin{bmatrix} 9 \\ 5 \end{bmatrix} = \begin{bmatrix} -3 \times 9 + 5 \times 5 \\ 2 \times 9 + -3 \times 5 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} = c$$

We know that $AA^{-1} = \mathbb{1}$ therefore:

$$Ac = b$$

$$A^{-1}Ac = A^{-1}b$$

$$c = A^{-1}b$$

Question 2.2

Derivatives

1.

$$f(x) = x^2 + 2x + 3$$

$$\frac{\partial f}{\partial x} = 2x + 2$$

2.

$$f(x) = (2x^3 + 1)^2 = 4x^6 + 4x^3 + 1$$

$$\frac{\partial f}{\partial x} = 24x^5 + 12x^2$$

Partial Derivatives

1. We apply the product rule of derivation

$$\textit{in general} \quad f(x) = g(x)e(x) \to f'(x) = g'(x)e(x) + g(x)e'(x)$$

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

$$\frac{\partial f}{\partial x} = 2(x + 2y)\sin(xy) + (x + 2y)^2 \cos(xy)y$$

$$\frac{\partial f}{\partial y} = 4(x + 2y)\sin(xy) + (x + 2y)^2 \cos(xy)x$$

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

2.

$$f(x) = log(g(x)) \to f'(x) = \frac{1}{g(x)} \cdot g'(x)$$

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

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

$$\frac{\partial f}{\partial y} = 2 \cdot \frac{1}{x + y^2 - z} \cdot 2y = \frac{4y}{x + y^2 - z}$$

$$\frac{\partial f}{\partial z} = 2 \cdot \frac{1}{x + y^2 - z} \cdot (-1) = \frac{-2}{x + y^2 - z}$$

3.

$$f(x) = \exp(g(x)) \to f'(x) = \exp(g(x)) \cdot g'(x)$$

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

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

$$\frac{\partial f}{\partial y} = \exp(x\cos(y+z)) \cdot -x\sin(y+z) = -x\exp(x\cos(y+z))\sin(y+z)$$

$$\frac{\partial f}{\partial z} = \exp(x\cos(y+z)) \cdot -x\sin(y+z) = -x\exp(x\cos(y+z))\sin(y+z)$$

Question 2.3

1.

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

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

$$= \mu^T(\Sigma^{-1}+S^{-1})\mu - \underbrace{\mu^T(\Sigma^{-1}x + S^{-1}\mu_0)}_{V(\mu)} - \underbrace{(x^T\Sigma^{-1} + \mu_0^T S^{-1})\mu}_{Q(\mu)} + x^T\Sigma^{-1}x + \mu_0^T S^{-1}\mu_0$$

We know that for $x \in \mathbb{R}$ and symmetric matrix $A \in \mathbb{R}^{D\times D}$

$$x = x^T; A = A^T$$

Therefore $V(\mu) = Q(\mu)$, as both $Q(\mu)$ and $V(\mu)$ result in scalars.

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

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

We can continue to gather our terms further.

$$= \mu^T(\Sigma^{-1}+S^{-1})\mu - 2\mu^T(\Sigma^{-1}x + S^{-1}\mu_0) + \underbrace{x^T\Sigma^{-1}x}_{M} + \underbrace{\mu_0^T S^{-1}\mu_0}_{N}$$

As M and N do not depend on μ, we can summarize both expressions to a constant c

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

If we define $E(\mu) = (\mu - (x+\mu_0))^T(\Sigma^{-1}+S^{-1})(\mu-(x+\mu_0))$, we find that:

$$E(\mu) = \mu^T(\Sigma^{-1}+S^{-1})\mu - \mu^T\Sigma^{-1}x + \underbrace{\mu^T\Sigma^{-1}\mu_0}_{Z_1} + \underbrace{\mu^T S^{-1}x}_{Z_3} + \mu^T S^{-1}\mu_0 - x^T\Sigma^{-1}\mu + \underbrace{x^T S^{-1}\mu}_{Z_4} +$$

$$\mu_0^T S^{-1}\mu + \underbrace{\mu_0^T\Sigma^{-1}\mu}_{Z_2} + \underbrace{(x+\mu_0)^T(\Sigma^{-1}+S^{-1})(x+\mu_0)}_{constant\ d}$$

Upon further simplification, we find that $F(\mu) = E(\mu) + d'$ as $Z_1 = Z_2$ and $Z_3 = Z_4$.

2.

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

3.

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

$$\text{We know } \frac{\partial}{\partial x} x^T A x = x^T(A + A^T)$$

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

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

$$\frac{\partial F(\mu)}{\partial \mu} = 0$$

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

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

Given that matrices Σ and S are invertible, it follows that:

$$\det(\Sigma^{-1}) > 0$$

$$\det(S^{-1}) > 0$$

$$\det(\Sigma) > 0$$

$$\det(S) > 0$$

Using Minowski's theorem, we show that the following sum is therefore invertible: $\Sigma^{-1} + S^{-1}$.

$$\det(A + B) \geq \det(A) + \det(B)$$

$$\det(\Sigma^{-1} + S^{-1}) \geq \det(\Sigma^{-1}) + \det(S^{-1}) > 0$$