SHRUTI RAO
2636454
s2.rao@student.vu.nl

Homework Assignment 1
Machine Learning 1, 19/20

2019-09-13

Multivariate Calculus

Question 2.1.a

$$
\begin{aligned}
\nabla_\mu (x-\mu)^T \sum{}^{-1}(x-\mu) &= (x^T - \mu^T)(\sum{}^{-1}x - \sum{}^{-1}\mu) \\
&= x^T \sum{}^{-1}x - x^T \sum{}^{-1}\mu - \mu^T \sum{}^{-1}x + \mu^T \sum{}^{-1}\mu
\end{aligned}
$$

Partially differentiating with respect to μ:

$$
\begin{aligned}
\frac{df}{d\mu} &= \frac{df(x^T \sum{}^{-1}x)}{d\mu} - \frac{df(x^T \sum{}^{-1}\mu)}{d\mu} - \frac{df(\mu^T \sum{}^{-1}x)}{d\mu} + \frac{df(\mu^T \sum{}^{-1}\mu)}{d\mu} \\
&= 0 - (x^T \sum{}^{-1}) - (\sum{}^{-1^T} x^T) + (\mu^T(\sum{}^{-1} + \sum{}^{-1^T})) && \text{since} \frac{df}{dX}(X^TAX) = X^T(A+A^T) \\
&= -(x^T \sum{}^{-1}) - (\sum{}^{-1}x^T) + (2\mu^T \sum{}^{-1}) && \text{transpose of inverse is inverse} \\
&= -2x^T \sum{}^{-1} + 2\mu^T \sum{}^{-1} \\
&= -2(x^T - \mu^T) \sum{}^{-1}
\end{aligned}
$$

Question 2.1.b

Assuming log to represent natural logarithm:

$$\nabla_q - p^T\ log(q) = -p^T (q)^T q^{-1} \qquad \text{since } ln(X) = X^T X^{-1}$$

Question 2.1.c

Given:
$\nabla_w f$
$f \ = \ Wx$
$W \in \mathbb{R}^{2x3}$
$x \in \mathbb{R}^3$
Following Example 5.12 we require the gradient of $\frac{df}{dw}$.
The determinant of the gradient is determined to be $\frac{df}{dw} \in^{2x(2x3)}$
Representing the gradient as a collection of partial derivatives:

$$\frac{df}{dw} = \begin{bmatrix} \frac{df_1}{dw} \\ \frac{df_2}{dw} \end{bmatrix} \qquad \frac{df_i}{dw} \in \mathbb{R}^{1x(2x3)}$$

The matrix vector multiplier is:

$$f_i = \sum_{j=1}^{3} w_{ij}x_j \qquad \text{where } i = 1, 2$$

The partial derivatives are then given as:

$$\frac{\partial f_i}{\partial w_{iq}} = x_q$$

The partial derivative of f_i with respect to a row of W:

$$\frac{\partial f_i}{\partial w_{iq,:}} = x^T \in \mathbb{R}^{1x1x3}$$
$$\frac{\partial f_i}{\partial w_{k\neq i}} = 0^T \in \mathbb{R}^{1x1x3}$$

Stacking the partial derivatives:

$$\begin{bmatrix} x^T \\ 0^T \\ 0^T \\ x^T \end{bmatrix} \in \mathbb{R}^{1x(2x3)}$$

Question 2.1.d

Given:
$\nabla_w f$
$f = (\mu - Wx)^T \sum{}^{-1}(\mu - Wx)$
$W \in \mathbb{R}^{MxK}$

$$\begin{aligned}(\mu - Wx)^T \sum{}^{-1}(\mu - Wx) &= (\mu^T - W^T x^T)(\sum{}^{-1}\mu - \sum{}^{-1}Wx) \\ &= \mu^T \sum{}^{-1}\mu - \mu^T \sum{}^{-1}Wx - W^T x^T \sum{}^{-1}\mu + W^T x^T \sum{}^{-1}Wx)\end{aligned}$$

Now partially differentiating:

$$\begin{aligned}&= \frac{\partial d(\mu^T \sum{}^{-1}\mu)}{dW} - \frac{\partial d(\mu^T \sum{}^{-1}Wx)}{dW} - \frac{\partial d(W^T x^T \sum{}^{-1}\mu)}{dW} + \frac{\partial d(W^T x^T \sum{}^{-1}Wx))}{dW} \\ &= 0 - (\mu^T \sum{}^{-1}x) - (x \sum{}^{-1}\mu^T) + W^T(x^T \sum{}^{-1}x + x\sum{}^{-1}x^T) \qquad \text{since } \frac{df}{dX}(X^T AX) = X^T(A + A^T) \\ &= -2(\mu^T \sum{}^{-1}x) + 2(W^T x^T \sum{}^{-1}x) \\ &= -2(\mu^T - W^T x^T)\sum{}^{-1}x\end{aligned}$$

Probability Theory

Question 3.1.a

The likelihood or degree of plausibility of a robbery, given that the man just climbed through the broken window of a jewellery store is quite high. This high likelihood is implied by the presence of conditions - *man running*, *a bag*, *a broken window*, *jewellery store* and *night time conditions*. Finally, previous knowledge of a man running at nighttime from a jewellery store with a bag being known as a criminal can further improve the likelihood of this resulting to be a criminal.

Question 3.1.b

The following notation can be formalized:
O where O is a random variable that determines if person made observation that it is a criminal
$C = \{c, nc\}$ where C is a random variable that determines if person is criminal or not; c indicates criminal; nc

indicates not criminal

Based on the given information:
probability of being a criminal: $P(C = c) = 1/10^5$
probability of not being a criminal: $P(C = nc) = 1 - 1/10^5$

Based on the problem statement and using the Bayesian Rule:

1. Probability of making observation that man is a criminal and man turns out to be a criminal:
$$\begin{aligned} p(O|C = c) &= \frac{p(C = c|O) * p(O)}{p(C = c)} \\ &= 0.8 \end{aligned}$$

2. Probability of making observation that man is a criminal when man is not a criminal:
$$\begin{aligned} p(O|C = nc) &= \frac{p(C = nc|O) * p(O)}{p(C = nc)} \\ &= \frac{1}{10^6} \end{aligned}$$

Question 3.1.c

Given information:
$p(O|C = c) = 0.8$ $p(C = c) = \frac{1}{10^5}$ $P(O) =$
We first need to find the probability for making an observation that the man is a criminal: $p(O)$. This can be calculated using the sum and product rules altogether:

$$\begin{aligned} p(O) &= p(O|C = c) * p(C = c) + p(O|C = nc) * p(C = nc) \\ &= 0.8 * \frac{1}{10^5} + \frac{1}{10^6} * (1 - \frac{1}{10^5}) \\ &= 0,000008 + 0,000001 \\ &= 0,000009 \ or \ \frac{9}{10 * 6} \end{aligned}$$

Next, we can calculate the probability of the man being a criminal:

$$\begin{aligned} p(C = c|O) &= \frac{p(O|C = c) * p(C = c)}{p(O)} \\ &= \frac{0.8 * \frac{1}{10^5}}{\frac{9}{10^6}} \\ &= 0,888888889 \end{aligned}$$

There is 89% probability that the man observed is indeed a criminal.

Question 3.1.d

The belief of original conditions (such as night time, man and jewellery store) are updated to realize that children and morning can be included in judging a robbery. While the two incidents are independent of one another, the probability value calculated in part (c) will more or less remain the same.

Question 3.2.a

The data is modelled with Multinomial Distribution where ρ_1, ρ_2, ρ_3 and ρ_4 each represent the probability of drawing a card from a suit.

$$\begin{aligned} p(\mathcal{D}|\mu) &= \prod_{x=i}^{4} p(x_i|\rho) \\ &= \prod Mult(\rho_1, \rho_2, \rho_3, \rho_4) \end{aligned}$$

Question 3.2.b

$$\begin{aligned} p(X_1 = 4, X_2 = 4) &= (\frac{N!}{X_1!\, X_2!}) * P^N * P^N \\ &= (\frac{8!}{4!\, 4!}) * 0.25^8 * 0.25^8 \\ &= 1.62981451e - 8 \end{aligned}$$

Question 3.2.c

$$p(x, \mu) = \prod_{k=1}^{K} \mu_k^{x_k}$$

Question 3.2.f

The general form would be:

$$\underbrace{p(p|\mathcal{D})}_{posterior} = \frac{\overbrace{p(\mathcal{D}|p)}^{likelihood}\ .\overbrace{p(p)}^{prior}}{\underbrace{p(\mathcal{D})}_{evidence}}$$