

Chapter 7

stevenjin8

October 3, 2020

1 Proofs

Equation 7.54

One of the results that I did not find trivial was equation 7.54. Perhaps it was proven somewhere in Chapter 4. I prove the in this subsection. As a reminder, the equation in question is

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \bar{y}\mathbf{1}_N - \mathbf{X}\mathbf{w}\|_2^2\right).$$

As the author says, we must "integrate $[\mu]$ out":

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2) = \int p(\mathbf{y}, \mu|\mathbf{X}, \mathbf{w}, \sigma^2)d\mu.$$

Since μ is independent and $p(\mu)$ is constant for all $\mu \in \mathbb{R}$,

$$\begin{aligned} p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2) &\propto \int p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \mu, \sigma^2)d\mu \\ &\propto \int \exp\left(\frac{(\mathbf{y} - \mu\mathbf{1}_N - \mathbf{X}\mathbf{w})^2}{-2\sigma^2}\right) d\mu. \end{aligned}$$

In this section, $\mathbf{v}^2 = \mathbf{v} \cdot \mathbf{v}$ for any vector \mathbf{v} .

Next, we expand and take all terms that are independent of μ out of the integral:

$$\begin{aligned} p(\mathbf{y}|\mathbf{X}, \mathbf{w}, \sigma^2) &\propto \\ &\int \exp\left(\frac{\mathbf{y}^2 - 2\mathbf{y} \cdot (\mu\mathbf{1}) + (\mu\mathbf{1})^2 + 2(\mu\mathbf{1}) \cdot (\mathbf{X}\mathbf{w}) + (\mathbf{X}\mathbf{w})^2 + 2(\mathbf{X}\mathbf{w}) \cdot \mathbf{y}}{-2\sigma^2}\right) d\mu \\ &\propto \exp\left(\frac{\mathbf{y}^2 + 2(\mathbf{X}\mathbf{w}) \cdot \mathbf{y} + (\mathbf{X}\mathbf{w})^2}{-2\sigma^2}\right) \\ &\int \exp\left(\frac{-2\mathbf{y} \cdot (\mu\mathbf{1}) + (\mu\mathbf{1})^2 + 2(\mu\mathbf{1}) \cdot (\mathbf{X}\mathbf{w})}{-2\sigma^2}\right) d\mu. \end{aligned}$$

Since $\sum_i x_{ij} = 0$, we can ignore $\mu \mathbf{1} \cdot \mathbf{X} \mathbf{w}$, and complete the square:

$$\begin{aligned}
& \int \exp \left(\frac{-2\mathbf{y} \cdot (\mu \mathbf{1}) + (\mu \mathbf{1})^2 + (\mu \mathbf{1}) \cdot (\mathbf{X} \mathbf{w})}{-2\sigma^2} \right) d\mu \\
&= \int \exp \left(\frac{N(-2\mu \bar{y} + \mu^2)}{-2\sigma^2} \right) d\mu \\
&= \exp \left(\frac{-N\bar{y}^2}{-2\sigma^2} \right) \int \exp \left(\frac{N(\mu - \bar{y})^2}{-2\sigma^2} \right) d\mu \\
&\propto \exp \left(\frac{-N\bar{y}^2}{-2\sigma^2} \right) \\
&= \exp \left(\frac{-(\bar{y} \mathbf{1})^2}{-2\sigma^2} \right)
\end{aligned} \tag{1}$$

Plugging this result back into equation (1) yields

$$\begin{aligned}
p(\mathbf{y} | \mathbf{X}, \mathbf{w}, \sigma^2) &\propto \exp \left(\frac{\mathbf{y}^2 - (\bar{y} \mathbf{1})^2 - 2(\mathbf{X} \mathbf{w}) \cdot \mathbf{y} + (\mathbf{X} \mathbf{w})^2}{-2\sigma^2} \right) \\
&= \exp \left(\frac{(\mathbf{y} - \bar{y} \mathbf{1})^2 - 2(\mathbf{X} \mathbf{w}) \cdot \mathbf{y} + (\mathbf{X} \mathbf{w})^2}{-2\sigma^2} \right) \\
&= \exp \left(\frac{(\mathbf{y} - \bar{y} \mathbf{1})^2 - 2(\mathbf{X} \mathbf{w}) \cdot (\mathbf{y} - \bar{y} \mathbf{1}) + (\mathbf{X} \mathbf{w})^2}{-2\sigma^2} \right) \\
&= \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \bar{y} \mathbf{1}_N - \mathbf{X} \mathbf{w}\|_2^2 \right).
\end{aligned} \tag{2}$$

Exercises

Exercise 9

In this exercise we use the results of section 4.3.1 to arrive at the same formula as that of exercise 7.5:

$$\mathbb{E}[y | \mathbf{x}] = \bar{y} - \mathbf{w}^T \bar{\mathbf{x}} + \mathbf{w}^T \mathbf{x}$$

(this formula is not the exact one given in the question but they are equivalent). First, we find the covariance matrices Σ_{XX} and Σ_{XY} of the joint distribution:

$$\begin{aligned}
\Sigma &= \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix} \\
&= (\mathbf{y} \quad \mathbf{X})^T (\mathbf{y} \quad \mathbf{X}) \\
&= \begin{pmatrix} \mathbf{y}^T \mathbf{y} & \mathbf{y}^T \mathbf{X} \\ \mathbf{X}^T \mathbf{y} & \mathbf{X}^T \mathbf{X} \end{pmatrix}.
\end{aligned}$$

Thus,

$$\begin{aligned}\Sigma_{XX} &= \mathbf{X}^T \mathbf{X} \\ \Sigma_{YX} &= \mathbf{y}^T \mathbf{X}.\end{aligned}\tag{3}$$

I know the question said to find Σ_{XY} , but I think we are supposed to find Σ_{YX} . Finding the means μ_x and μ_y is a lot easier:

$$\begin{aligned}\mu_x = \bar{\mathbf{x}} &= \frac{1}{N} \sum \mathbf{x}_i \\ \mu_y = \bar{y} &= \frac{1}{N} \sum y_i\end{aligned}\tag{4}$$

Now, we plug (3) and (4) into equation 4.69 and replace 1 with y and 2 with x . Given \mathbf{x} our prediction for y is

$$\begin{aligned}\mu_{y|x} &= \mu_y + \Sigma_{YX} \Sigma_{XX}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \bar{y} + \mathbf{y} \mathbf{X}^T (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \bar{y} + ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y})^T (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \bar{y} + -\mathbf{w}^T \bar{\mathbf{x}} + \mathbf{w}^T \mathbf{x} \\ &= \mathbb{E}[y|\mathbf{x}].\end{aligned}$$

Recall that $\mathbf{X}^T \mathbf{X}$ is symmetric so $(\mathbf{X}^T \mathbf{X})^{-T} = (\mathbf{X}^T \mathbf{X})^{-1}$. I find it reassuring that the discriminative and generative approach converge to the same solution. That being said, I am not sure what the author is looking for in part b) as in both approaches, one ends up doing the same calculations.