# Introduction to ML: HW1 solution

**Q1) Derive equation 2.17. (hint: use equation 2.16)**

$$w_1 = \frac{\sum_t r^t x^t - \overline{x}\overline{r}N}{\sum_t (x^t)^2 - N\overline{x}^2}$$

We take the derivative of the sum of squared errors with respect to the two parameters, set them equal to 0, and solve these two equations in two unknowns:

$$\begin{aligned}
E(w_1, w_0|\mathcal{X}) &= \frac{1}{N}\sum_{i=1}^{N}\left[r^t - (w_1 x^t + w_0)\right]^2 \\
\frac{\partial E}{\partial w_0} &= \sum_t \left[r^t - (w_1 x^t + w_0)\right] = 0 \\
\sum_t r^t &= w_1 \sum_t x^t + N w_0 \\
w_0 &= \sum_t r^t/N - w_1 \sum_t x^t/N = \overline{r} - w_1\overline{x} \\
\frac{\partial E}{\partial w_1} &= \sum_t \left[r^t - (w_1 x^t + w_0)\right] x^t = 0 \\
\sum_t r^t x^t &= w_1 \sum_t (x^t)^2 + w_0 \sum_t x^t \\
\sum_t r^t x^t &= w_1 \sum_t (x^t)^2 + (\overline{r} - w_1\overline{x}) \sum_t x^t \\
\sum_t r^t x^t &= w_1 \left(\sum_t (x^t)^2 - \overline{x}\sum_t x^t\right) + \overline{r}\sum_t x^t \\
\sum_t r^t x^t &= w_1 \left(\sum_t (x^t)^2 - \overline{x}N\overline{x}\right) + \overline{r}N\overline{x} \\
w_1 &= \frac{\sum_t r^t x^t - \overline{x}\overline{r}N}{\sum_t (x^t)^2 - N\overline{x}^2}
\end{aligned}$$

**Q2) Assume a disease so rare that it is seen in only one person out of every million. Assume also that we have a test that is effective in that if a person has the disease, there is a 99 percent chance that the test result will be positive; however, the test is not perfect, and there is a one in a thousand chance that the test result will be positive on a healthy person. Assume that a new patient arrives and the test result is positive. What is the probability that the patient has the disease?**

Let us represent disease by $d$ and test result by $t$. We are given the following: $P(d = 1) = 10^{-6}, P(t = 1|d = 1) = 0.99, P(t = 1|d = 0) = 10^{-3}$. We are asked $P(d = 1|t = 1)$.
We use Bayes' rule:

$$\begin{aligned}
P(d = 1|t = 1) &= \frac{P(t = 1|d = 1)P(d = 1)}{P(t = 1)} \\
&= \frac{P(t = 1|d = 1)P(d = 1)}{P(t = 1|d = 1)P(d = 1) + P(t = 1|d = 0)P(d = 0)} \\
&= \frac{0.99 \cdot 10^{-6}}{0.99 \cdot 10^{-6} + 10^{-3} \cdot (1 - 10^{-6})} = 0.00098902
\end{aligned}$$

That is, knowing that the test result is positive increased the probability of disease from one in a million to one in a thousand.

**Q3) Show that as we move an item from the consequent to the antecedent, confidence can never increase: confidence (ABC → D) ≥ confidence (AB → CD).**

$$\begin{aligned}
\text{confidence}(ABC \to D) &= \cancel{\frac{P(A, B, C, D)}{P(D)}} \\
\text{confidence}(AB \to CD) &= \cancel{\frac{P(A, B, C, D)}{P(C, D)}}
\end{aligned}$$

Since $P(C, D) = P(C|D)P(D) \leq P(D)$ because $P(C|D) \leq 1$, we have $\text{confidence}(ABC \to D) \geq \text{confidence}(AB \to CD)$.

**Q4) Let us say that we toss a fair coin ten times. What is the probability that we see four to six heads? What is the probability that we see forty to sixty heads when we toss it one hundred times?**

Let $p_0$ be the probability of heads. The probability that we see $x$ heads in $n$ tosses is given by the binomial distribution

$$P(x|n, p_0) = \binom{n}{x} p_0^x (1 - p_0)^{n-x}$$

The probability that we see four to six heads when we toss a fair coin ten times is

$$P(4 \le x \le 6 \,|\, 10, 1/2) = \sum_{x=4}^{6} \binom{10}{x} 1/2^x (1 - 1/2)^{n-x}$$

The probability that we see forty to sixty heads when we toss a fair coin one hundred times is

$$P(40 \le x \le 60 \,|\, 100, 1/2) = \sum_{x=40}^{60} \binom{100}{x} 1/2^x (1 - 1/2)^{n-x}$$

which is a larger value.

**Q5) Write the log likelihood for a multinomial sample and show equation 4.6.**

$$p_i = \frac{\sum_t x_i^t}{N}$$

We add the constraint as a Lagrange term and maximize it:

$$\begin{aligned}
J(p_i) &= \sum_i \sum_t x_i^t \log p_i + \lambda(1 - \sum_i p_i) \\
\frac{\partial J}{\partial p_i} &= \frac{\sum_t x_i^t}{p_i} - \lambda = 0 \\
\lambda &= \frac{\sum_t x_i^t}{p_i} \Rightarrow p_i \lambda = \sum_t x_i^t \\
\sum_i p_i \lambda &= \sum_i \sum_t x_i^t \Rightarrow \lambda = \sum_t \sum_i x_i^t \\
p_i &= \frac{\sum_t x_i^t}{\sum_t \sum_i x_i^t} = \frac{\sum_t x_i^t}{N} \text{ because } \sum_i x_i^t = 1
\end{aligned}$$

**Q6) Show equation 5.11.**

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)\right]$$

Given that

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

we have

$$\begin{aligned} |\mathbf{\Sigma}| &= \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2 = \sigma_1^2\sigma_2(1-\rho^2) \\ |\mathbf{\Sigma}|^{1/2} &= \sigma_1\sigma_2\sqrt{1-\rho^2} \\ \mathbf{\Sigma}^{-1} &= \frac{1}{\sigma_1^2\sigma_2(1-\rho^2)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \end{aligned}$$

and $(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ can be expanded as

$$[x_1 - \mu_1 \; x_2 - \mu_2]\begin{bmatrix} \frac{\sigma_2^2}{\sigma_1^2\sigma_2(1-\rho^2)} & -\frac{\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2(1-\rho^2)} \\ -\frac{\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2(1-\rho^2)} & \frac{\sigma_1^2}{\sigma_1^2\sigma_2(1-\rho^2)} \end{bmatrix}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

$$= \frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$