MAD 2019/2020 Exam Answers

Exam Number: 18

January 13, 2020

Exercise 1

I find the task of this exercise a little ambiguous, as I am not sure if I am supposed to (only) compute the mean point of the data, the two eigenvalues and the two eigenvectors or if I am also supposed to convert the dimensionality. To be on the safe side, I will also convert the dimensionality, however, if I was not supposed to do this, please just overlook that part of my answer.

We are given the following $N = 8$ 2-dimensional points:

coordinate x	0.1	0.5	1.1	-0.5	1.3	0.2	-0.1	1
coordinate y	1	1	2	0.2	-0.1	-0.1	-1.5	-2.5

First of we need to make each row have zero mean. This is done, by first computing $\bar{x}$ and $\bar{y}$:

$$\bar{x} = \frac{1}{8} \sum_{n=1}^{8} x_n = 0.45, \qquad \bar{y} = \frac{1}{8} \sum_{n=1}^{8} y_n = 0$$

Which gives us the mean point of data, $(0.45, 0)$. $\bar{x}$ and $\bar{y}$ are then subtracted from respectively coordinate x and coordinate y, which results in the following rows, with zero mean:

coordinate x	-0.35	0.05	0.65	-0.95	0.85	-0.25	-0.55	0.55
coordinate y	1	1	2	0.2	-0.1	-0.1	-1.5	-2.5

From this we can extract the following matrix:

$$S = \begin{bmatrix} -0.35 & 1 \\ 0.05 & 1 \\ 0.65 & 2 \\ -0.95 & 0.2 \\ 0.85 & -0.1 \\ -0.25 & -0.1 \\ -0.55 & -1.5 \\ 0.55 & -2.5 \end{bmatrix}$$

Which we use to compute the covariance matric C:

$$C = \frac{1}{N} \cdot S^T \cdot S =$$

$$\frac{1}{8} \cdot \begin{bmatrix} -0.35 & 0.05 & 0.65 & -0.95 & 0.85 & -0.25 & -0.55 & 0.55 \\ 1 & 1 & 2 & 0.2 & -0.1 & -0.1 & -1.5 & -2.5 \end{bmatrix} \cdot \begin{bmatrix} -0.35 & 1 \\ 0.05 & 1 \\ 0.65 & 2 \\ -0.95 & 0.2 \\ 0.85 & -0.1 \\ -0.25 & -0.1 \\ -0.55 & -1.5 \\ 0.55 & -2.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.355 & 0.025 \\ 0.025 & 1.82 \end{bmatrix}$$

1

Now that we have found the covariance matrix $\boldsymbol{C}$, we can find the eigenvalues. First we need to compute $\boldsymbol{C} - \lambda \cdot \boldsymbol{I}_2$:

$$\boldsymbol{C} - \lambda \cdot \boldsymbol{I}_2 = \begin{bmatrix} 0.355 & 0.025 \\ 0.025 & 1.82 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 0.355 - \lambda & 0.025 \\ 0.025 & 1.82 - \lambda \end{bmatrix}$$

Which we need to find the determinant of:

$$det(\boldsymbol{C} - \lambda \cdot \boldsymbol{I}_2) = \begin{vmatrix} 0.355 - \lambda & 0.025 \\ 0.025 & 1.82 - \lambda \end{vmatrix}$$
$$= (0.355 - \lambda) \cdot (1.82 - \lambda) - 0.025 * 0.025$$
$$= 0.355 \cdot 1.82 - \lambda \cdot 1.82 + \lambda^2 - 0.025 \cdot 0.025 - 0.355 \cdot \lambda$$
$$\approx \lambda^2 - 2.175\lambda + 0.6455$$

And now solve for λ:
$$\lambda^2 - 2.175\lambda + 0.6455 = 0$$

$\Leftrightarrow$

$$\lambda \, \frac{-(-2.175) \pm \sqrt{(-2.175)^2 - 4 \cdot 1 \cdot 0.6455}}{2 \cdot 1} \approx \begin{cases} 1.820 \\ 0.355 \end{cases}$$

Thus, we have found the two eigenvalues $\lambda_1 = 1.820$ and $\lambda_2 = 0.355$.

Now, for finding the associated eigenvectors, we first let $\lambda = \lambda_1$ and find $\boldsymbol{B} = \boldsymbol{C} - \lambda \cdot \boldsymbol{I}_2$:

$$\boldsymbol{B} = \boldsymbol{C} - \lambda \cdot \boldsymbol{I}_2 = \begin{bmatrix} 0.355 - 1.820 & 0.025 \\ 0.025 & 1.82 - 1.820 \end{bmatrix}$$
$$= \begin{bmatrix} -1.4658 & 0.025 \\ 0.025 & 0 \end{bmatrix}$$

Now, we put $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$ on row echelon form

$$\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$$

$\Rightarrow$

$$\left[\begin{array}{cc|c} -1.4658 & 0.025 & 0 \\ 0.025 & 0 & 0 \end{array} \right]$$

Which is done by adding row 1 times 0.0171 to row 2. Thus, we get the following augmented matrix:

$$\left[\begin{array}{cc|c} -1.4658 & 0.025 & 0 \\ 0 & 0 & 0 \end{array} \right]$$

Thus, we have the following equation, where we isolate x_1:

$$-1.4658 \cdot x_1 + 0.025 \cdot x_2 = 0$$

$\Leftrightarrow$

$$0.025 \cdot x_2 = 1.4658 \cdot x_1$$

2

$\Leftrightarrow$

$$0.171 \cdot x_2 = x_1$$

From this point, we can choosing any integer for x_2 and get the eigenvector for the eigenvalue 1.820. I am here choosing $x_2 = 1$:

$$0.171 \cdot 1 = x_1$$

$\Leftrightarrow$

$$0.171 = x_1$$

Thus, the eigenvector for the eigenvalue 1.820 is:

$$\begin{bmatrix} 0.171 \\ 1 \end{bmatrix}$$

Now, we let $\lambda = \lambda_2$ and follow the same procedure:

$$\boldsymbol{B} = \boldsymbol{C} - \lambda \cdot \boldsymbol{I}_2 = \begin{bmatrix} 0.355 - 0.355 & 0.025 \\ 0.025 & 1.82 - 0.355 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0.025 \\ 0.025 & 1.465 \end{bmatrix}$$

$$\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$$

$\Rightarrow$

$$\left[\begin{array}{cc|c} 0 & 0.025 & 0 \\ 0.025 & 1.465 & 0 \end{array} \right]$$

To put the matrix on row echelon form, we simply swap row 1 and row 2 and optain the following matrix

$$\left[\begin{array}{cc|c} 0.025 & 1.465 & 0 \\ 0 & 0.025 & 0 \end{array} \right]$$

Thus, we have the following equation, where x_1 needs to be isolated:

$$0.025 \cdot x_1 + 1.465 \cdot x_2 = 0$$

$\Leftrightarrow$

$$1.465 \cdot x_2 = -0.025 \cdot x_1$$

$\Leftrightarrow$

$$-58.6 \cdot x_2 = x_1$$

Again, here it is possible to choose any value for x_2, I will let $x_2 = 1$ and get the following:

$$-58.6 \cdot 1 = x_1$$

$\Leftrightarrow$

$$-58.6 = x_1$$

Thus, for the eigenvalue 0.355, we have the eigenvector

$$\begin{bmatrix} -58.6 \\ 1 \end{bmatrix}$$

Thus, sorting the eigenvalues in descending order and the eigenvectors accordingly, we get the following matrix of the eigenvectors:

$$\boldsymbol{P} = \begin{bmatrix} 0.171 & -58.6 \\ 1 & 1 \end{bmatrix}$$

The dimensionality of the given data can now be converted by multiplying the zero-mean matrix, $\boldsymbol{S}$ with $\boldsymbol{P}$:

$$\begin{bmatrix} -0.35 & 1 \\ 0.05 & 1 \\ 0.65 & 2 \\ -0.95 & 0.2 \\ 0.85 & -0.1 \\ -0.25 & -0.1 \\ -0.55 & -1.5 \\ 0.55 & -2.5 \end{bmatrix} \cdot \begin{bmatrix} 0.171 & -58.6 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.9402 & 21.51 \\ 1.0086 & -1.93 \\ 2.1112 & -36.09 \\ 0.03755 & 55.87 \\ 0.04535 & -49.91 \\ -0.1428 & 14.55 \\ -1.5941 & 30.73 \\ -2.406 & -34.73 \end{bmatrix}$$

Exercise 2

First, we compute $Gini(T_{org}, \lambda)$:

$$Gini(T_{org}, \lambda) = \frac{17}{28} \cdot gini(T_{first}) + \frac{11}{28} \cdot gini(T_{second})$$

$$= \frac{17}{28}\left(1 - \left(\frac{6}{17}\right)^2 - \left(\frac{1}{17}\right)^2 - \left(\frac{10}{17}\right)^2\right) + \frac{11}{28}\left(1 - \left(\frac{1}{11}\right)^2 - \left(\frac{7}{11}\right)^2 - \left(\frac{3}{11}\right)^2\right)$$

$$= 0.5206$$

Next, we compute $Gain(T_{orig}, \lambda)$:

$$Gain(T_{orig}, \lambda) = H(7, 8, 13) - \frac{17}{28} \cdot H(6, 1, 10) - \frac{11}{28} \cdot H(1, 7, 3)$$

$$= \left(-\frac{13}{28} \cdot \log_2\left(\frac{13}{28}\right) - \frac{8}{28} \cdot \log_2\left(\frac{8}{28}\right) - \frac{7}{28} \cdot \log_2\left(\frac{7}{28}\right)\right)$$

$$- \frac{17}{28}\left(-\frac{10}{17} \cdot \log_2\left(\frac{10}{17}\right) - \frac{6}{17} \cdot \log_2\left(\frac{6}{17}\right) - \frac{1}{17} \cdot \log_2\left(\frac{1}{17}\right)\right)$$

$$- \frac{11}{28}\left(-\frac{7}{11} \cdot \log_2\left(\frac{7}{11}\right) - \frac{3}{11} \cdot \log_2\left(\frac{3}{11}\right) - \frac{1}{11} \cdot \log_2\left(\frac{1}{11}\right)\right)$$
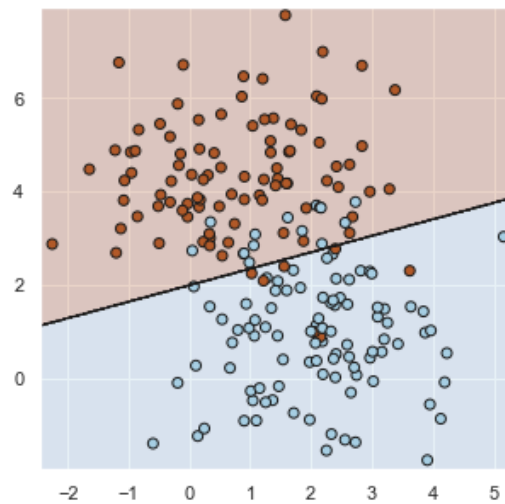
$$= 0.3015$$

Thus, it has been computed, that $Gini(T_{org}, \lambda) = 0.5206$ and $Gain(T_{orig}, \lambda) = 0.3015$

Exercise 3

With my implementation, I have the following performance of the cross-validation:

C	Accuracy
0.001	0.9225
0.01	0.92125
0.1	0.9274999999999999
1	0.9237500000000001
10	0.9237500000000001
100	0.9237500000000001

And the following decision boundary for the optimal svm classifier:
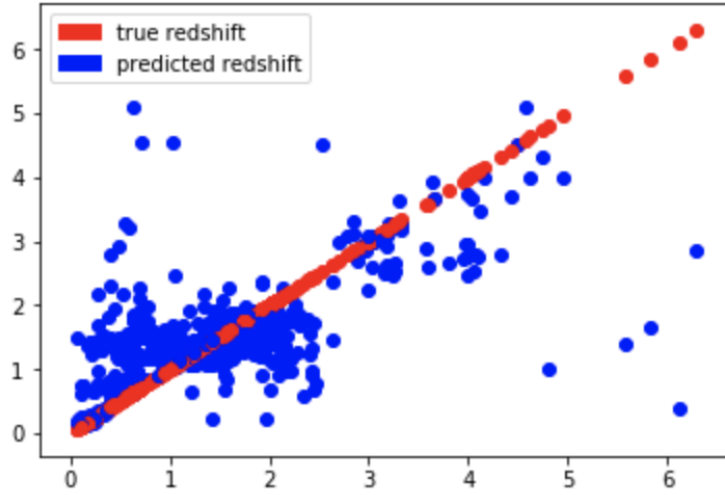

Exercise 4

Part 1

(a)

With my implementation, I have gotten the following RMSE-value:

$$RMSE = 0.8243064553494787$$
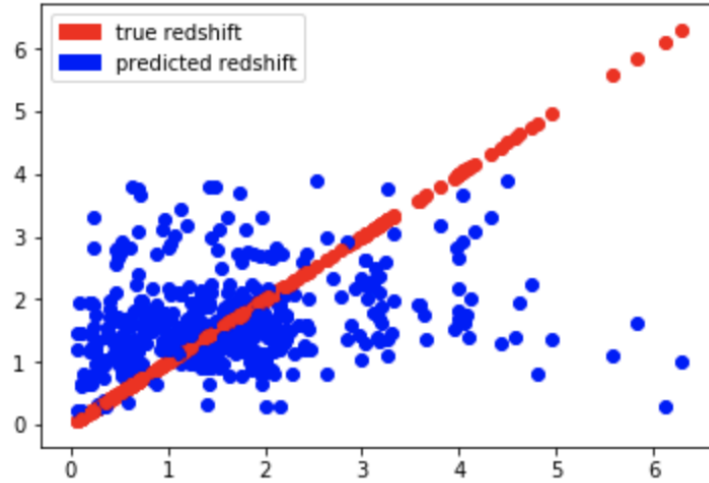
And the following scatter plot:

(b)

With this new distance formula, I have optained the following RMSE-value:

$$RMSE = 1.0997971796682453$$

And the following scatter plot:

The matrix M, makes it possible to tune the impact that the different features have on the distance between the two data point, thus also the impact on the prediction. For instance, the given example matrix

$$\begin{bmatrix} 0.00001 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.00001 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00001 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.00001 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00001 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.00001 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.00001 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.00001 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 \end{bmatrix}$$

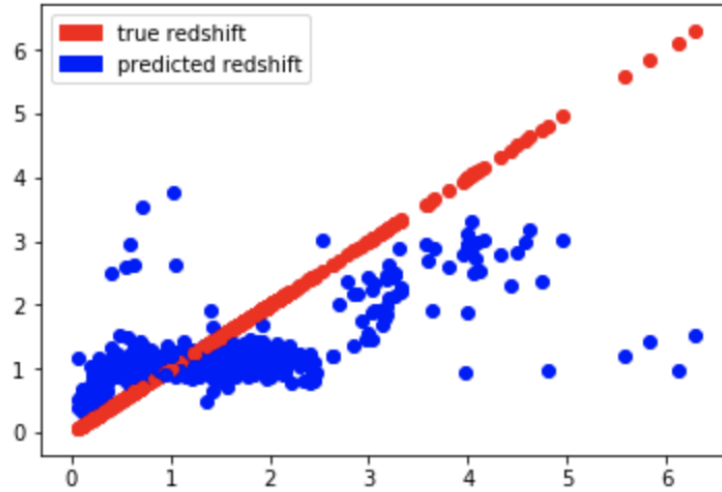makes the two last features have a bigger impact on the distance and prediction, than the first eight features. This could imply, that the two last feature are more "important", than the first eight features.

Part 2

With my implementation, I got the following RMSE-value:

$$RMSE = 0.8925765347348245$$

And the following scatter plot:

Exercise 5

Part 1

We have been given the following expression

$$p(t_n|\boldsymbol{x}_n; \boldsymbol{w}) = \frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!}$$

And we wish to find the joint conditional density:

$$p(\boldsymbol{t}|\boldsymbol{X}; \boldsymbol{w})$$

Which extends to

$$p(t_1, ..., t_N|\boldsymbol{x}_1, ..., \boldsymbol{x}_N; \boldsymbol{w})$$

Following the formula for joint probability distribution

$$p(A, B) = p(A) \cdot p(B)$$

We get the result

$$p(\boldsymbol{t}|\boldsymbol{X}; \boldsymbol{w}) = \prod_{n=1}^{N} p(t_n|\boldsymbol{x}_n; \boldsymbol{w}) = \prod_{n=1}^{N} \frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!}$$

Part 2

Since, it says in the exercise desciption, how we only should include the essential steps for deriving the expression, some of the descriptions of actions in the following exercise have been left out.

We have been given the expression for $p(w|\boldsymbol{t}, \boldsymbol{X}, \mu_0, \sigma_0^2)$, which we extend

$$p(w|\boldsymbol{t}, \boldsymbol{X}, \mu_0, \sigma_0^2) =$$
$$\frac{\prod_{n=1}^{N} \left(\frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!} \right) \frac{1}{(2\pi)^{\frac{D+1}{2}} \det(\sigma_0^2 \boldsymbol{I})^{\frac{1}{2}}} \exp(-\frac{1}{2}(\boldsymbol{w}-\mu_0)^T (\sigma_0^2 \boldsymbol{I})^{-1}(\boldsymbol{w}-\mu_0))}{p(\boldsymbol{t}|\boldsymbol{X})}$$

To find $\log p(w|\boldsymbol{t}, \boldsymbol{X}, \mu_0, \sigma_0^2)$, we start of by taking the log on both sides and reducing a little:

$$\log p(w|\boldsymbol{t}, \boldsymbol{X}, \mu_0, \sigma_0^2) =$$
$$\log \left(\frac{\prod_{n=1}^{N} \left(\frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!} \right) \frac{1}{(2\pi)^{\frac{D+1}{2}} \det(\sigma_0^2 \boldsymbol{I})^{\frac{1}{2}}} \exp(-\frac{1}{2}(\boldsymbol{w}-\mu_0)^T (\sigma_0^2 \boldsymbol{I})^{-1}(\boldsymbol{w}-\mu_0))}{p(\boldsymbol{t}|\boldsymbol{X})} \right)$$
$$= \log \left(\prod_{n=1}^{N} \left(\frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!} \right) \right) + \log \left(\frac{1}{(2\pi)^{\frac{D+1}{2}} \det(\sigma_0^2 \boldsymbol{I})^{\frac{1}{2}}} \right) +$$
$$\log \left(\exp(-\frac{1}{2}(\boldsymbol{w} - \mu_0)^T (\sigma_0^2 \boldsymbol{I})^{-1}(\boldsymbol{w} - \mu_0)) \right) - \log \left(p(\boldsymbol{t}|\boldsymbol{X}) \right)$$

For the sake of readability, I choose to split the proof into three sections:

First Section

In this section, I will prove

$$\log \left(\prod_{n=1}^{N} \left(\frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!} \right) \right) =$$
$$\sum_{n=1}^{N} \left(t_n \log f(\boldsymbol{x}_n; \boldsymbol{w}) - f(\boldsymbol{x}_n; \boldsymbol{w}) - \log t_n! \right)$$

This is done by the following

$$\log \left(\prod_{n=1}^{N} \left(\frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!} \right) \right) = \sum_{n=1}^{N} \log \left(\frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!} \right) =$$
$$\sum_{n=1}^{N} \left(\log(f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))) - \log t_n! \right) =$$
$$\sum_{n=1}^{N} \left(\log f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} + \log \exp(-f(\boldsymbol{x}_n;\boldsymbol{w})) - \log t_n! \right) =$$
$$\sum_{n=1}^{N} \left(t_n \log f(\boldsymbol{x}_n;\boldsymbol{w}) - f(\boldsymbol{x}_n;\boldsymbol{w}) - \log t_n! \right)$$

Second Section

In this section, I will prove

$$\log \left(\exp(-\tfrac{1}{2}(\boldsymbol{w} - \mu_0)^T (\sigma_0^2 \boldsymbol{I})^{-1} (\boldsymbol{w} - \mu_0)) \right) = -\tfrac{1}{2\sigma_0^2} (\boldsymbol{w} - \mu_0)^T (\boldsymbol{w} - \mu_0)$$

This is done by

$$\log \left(\exp(-\tfrac{1}{2}(\boldsymbol{w} - \mu_0)^T (\sigma_0^2 \boldsymbol{I})^{-1} (\boldsymbol{w} - \mu_0)) \right) = -\tfrac{1}{2}(\boldsymbol{w} - \mu_0)^T (\sigma_0^2 \boldsymbol{I})^{-1} (\boldsymbol{w} - \mu_0) =$$
$$-\tfrac{1}{2}(\boldsymbol{w} - \mu_0)^T (\tfrac{1}{\sigma_0^2} \boldsymbol{I})(\boldsymbol{w} - \mu_0) = -\tfrac{1}{2} \tfrac{1}{\sigma_0^2} (\boldsymbol{w} - \mu_0)^T \boldsymbol{I}(\boldsymbol{w} - \mu_0) =$$
$$-\tfrac{1}{2\sigma_0^2} (\boldsymbol{w} - \mu_0)^T (\boldsymbol{w} - \mu_0)$$

Third section

In this section, I will prove

$$\log \left(\frac{1}{(2\pi)^{\frac{D+1}{2}} \det(\sigma_0^2 \boldsymbol{I})^{\frac{1}{2}}} \right) = -\log \left(\sqrt{2\pi \sigma_0^2}^{D+1} \right)$$

This is done by

$$\log \left(\frac{1}{(2\pi)^{\frac{D+1}{2}} \det(\sigma_0^2 \boldsymbol{I})^{\frac{1}{2}}} \right) = \log(1) - \log((2\pi)^{\frac{D+1}{2}} \det(\sigma_0^2 \boldsymbol{I})^{\frac{1}{2}}) =$$
$$-\log((2\pi)^{\frac{D+1}{2}} \left(\sigma_0^{2 \cdot (D+1)} \right)^{\frac{1}{2}}) = -\log((2\pi)^{\frac{D+1}{2}} \left(\sigma_0^2 \right)^{\frac{D+1}{2}}) =$$
$$-\log((2\pi)^{\left(\frac{1}{2} \right)^D} \left(\sigma_0^2 \right)^{\left(\frac{1}{2} \right)^D}) = -\log(2\pi\sigma_0^2)^{\left(\frac{1}{2} \right)^D} = -\log \sqrt{(2\pi\sigma_0^2)}^D$$

Section 4

Substituting the results of these three sections into $\log p(w|\boldsymbol{t}, \boldsymbol{X}, \mu_0, \sigma_0^2)$ gives the following

$$\log p(w|\boldsymbol{t}, \boldsymbol{X}, \mu_0, \sigma_0^2) = \log \left(\prod_{n=1}^{N} \left(\frac{f(\boldsymbol{x}_n;\boldsymbol{w})^{t_n} \exp(-f(\boldsymbol{x}_n;\boldsymbol{w}))}{t_n!} \right) \right) +$$
$$\log \left(\frac{1}{(2\pi)^{\frac{D+1}{2}} \det(\sigma_0^2 \boldsymbol{I})^{\frac{1}{2}}} \right) + \log \left(\exp(-\tfrac{1}{2}(\boldsymbol{w} - \mu_0)^T (\sigma_0^2 \boldsymbol{I})^{-1} (\boldsymbol{w} - \mu_0)) \right) -$$
$$\log (p(\boldsymbol{t}|\boldsymbol{X})) = \sum_{n=1}^{N} \left(t_n \log f(\boldsymbol{x}_n;\boldsymbol{w}) - f(\boldsymbol{x}_n;\boldsymbol{w}) - \log t_n! \right) - \log \sqrt{(2\pi\sigma_0^2)}^D -$$
$$\tfrac{1}{2\sigma_0^2} (\boldsymbol{w} - \mu_0)^T (\boldsymbol{w} - \mu_0) - \log (p(\boldsymbol{t}|\boldsymbol{X}))$$

Rearranging, the equation a little, and we get to the result:

$$\sum_{n=1}^{N} \left(t_n \log f(\boldsymbol{x}_n; \boldsymbol{w}) - f(\boldsymbol{x}_n; \boldsymbol{w}) - \log t_n! \right) - \log \sqrt{(2\pi\sigma_0^2)}^D - \frac{1}{2\sigma_0^2} (\boldsymbol{w} -$$

$$\mu_0)^T (\boldsymbol{w} - \mu_0) - \log \left(p(\boldsymbol{t}|\boldsymbol{X}) \right)$$

$$= \sum_{n=1}^{N} \left(t_n \log f(\boldsymbol{x}_n; \boldsymbol{w}) - f(\boldsymbol{x}_n; \boldsymbol{w}) - \log t_n! \right) - \frac{1}{2\sigma_0^2} (\boldsymbol{w} - \mu_0)^T (\boldsymbol{w} - \mu_0) -$$

$$\log \sqrt{(2\pi\sigma_0^2)}^D - \log \left(p(\boldsymbol{t}|\boldsymbol{X}) \right), \qquad Q.E.D$$