Model Solution Tutorial 1

1.1. Moore - Penrose Pseudoinverse

Compute the pseudoinverse of

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

Begin by calculating the singular value decomposition

$$A \cdot A^T = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$$

$$p(\lambda) = \left\| \begin{matrix} 2-\lambda & 4 \\ 4 & 8-\lambda \end{matrix} \right\|$$

$$= (2-\lambda)(8-\lambda) - 16$$

$$= 16 - 10\lambda + \lambda^2 - 16$$

$$= \lambda^2 - 10\lambda = \lambda(\lambda - 10)$$

$$\Rightarrow \lambda_1 = 10 \quad , \lambda_2 = 0$$

Calculating the eigenpairs

$$\lambda_1 : \quad \begin{matrix} -8x_1 + 4x_2 = 0 \\ 4x_1 - 2x_2 = 0 \end{matrix} \quad \Rightarrow \text{normalized eigenvector } \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\lambda_2 : \quad \begin{matrix} 2x_1 + 4x_2 = 0 \\ 4x_1 + 8x_2 = 0 \end{matrix} \quad \Rightarrow \text{normalized eigenvector } \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

$$\Rightarrow \Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix}$$

$$U = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$A^T \cdot A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ 5 & 5 \end{pmatrix}$$

$$\Rightarrow p_2(\lambda) = \begin{Vmatrix} 5-\lambda & 5 \\ 5 & 5-\lambda \end{Vmatrix}$$

$$= (5-\lambda)^2 - 25$$

$$= \lambda^2 - 10\lambda = \lambda(\lambda - 10)$$

$$\Rightarrow \lambda_1 = 10, \quad \lambda_2 = 0$$

Calculating the eigenpairs

λ_1 : $\begin{aligned} -5x_1 + 5x_2 &= 0 \\ 5x_1 \cancel{+} - 5x_2 &= 0 \end{aligned}$ $\Rightarrow$ normalized eigenvector $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

λ_2 : $\begin{aligned} 5x_1 + 5x_2 &= 0 \\ 5x_1 + 5x_2 &= 0 \end{aligned}$ $\Rightarrow$ normalized eigenvector $\frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ +1 \end{pmatrix}$

$$\Rightarrow V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

The singular value decomposition is hence

$$\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}}_{U} \cdot \underbrace{\begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & +\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{V^T}$$

our pseudoinverse is then given by

$$A^+ = V \Sigma^+ U^*$$

$$\Rightarrow A^{+} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{10}} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \frac{1}{10} \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

1.2 Singular Value Decomposition

Compute the singular value decomposition of the matrix

b)

$$\begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$A \cdot A^{T} = \begin{pmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 & 0 \\ 4 & 3 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \overset{20}{\cancel{16}} & 14 & 0 & 0 \\ 14 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$p(\lambda) = \begin{vmatrix} \overset{20}{\cancel{16}} - \lambda & 14 & 0 & 0 \\ 14 & 10-\lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{vmatrix}$$

$$= (\overset{20}{\cancel{16}} - \lambda) \cdot \begin{vmatrix} 10-\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} - 14 \cdot \begin{vmatrix} 14 & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix}$$

$$= (\overset{20}{\cancel{16}} - \lambda)(10-\lambda)(\lambda^2) - 14 \cdot 14 \cdot \lambda^2$$

$$= \overset{200}{\cancel{160}} \lambda^2 - \overset{30}{\cancel{26}} \lambda^3 + \lambda^4 - 196 \lambda^2$$

$$= \lambda^2 (\lambda^2 - 26 \lambda - 36)$$

$$= \lambda^4 - 30\lambda^3 + 4\lambda^2$$

$$= \lambda^2(\lambda^2 - 30\lambda + 4)$$

$$\Rightarrow \lambda_1 = 15 + \sqrt{221}, \quad \lambda_2 = 15 - \sqrt{221}, \quad \lambda_{3,4} = 0$$

Calculating the eigenvectors

$$\lambda_1: \quad \left.\begin{aligned} (\overset{5}{\cancel{20-15}} - \sqrt{221})x_1 + 14x_2 &= 0 \\ 14x_1 + (\overset{-5}{\cancel{10-15}} - \sqrt{221})x_2 &= 0 \\ x_3 &= 0 \\ x_4 &= 0 \end{aligned}\right\}$$

$$x_3 = x_4 = 0 \quad x_1 = \frac{5+\sqrt{221}}{14}x_2$$

$$\lambda_2: \quad \left.\begin{aligned} (20 - 15 + \sqrt{221})x_1 + 14x_2 &= 0 \\ 14x_1 + (10 - 15 + \sqrt{221})x_2 &= 0 \\ x_3 &= 0 \\ x_4 &= 0 \end{aligned}\right\}$$

$$\Rightarrow x_3 = x_4 = 0 \quad \cancel{x_1 = 1}, \quad x_2 = -\frac{5+\sqrt{221}}{14}x_1$$

U is hence

$$U \approx \begin{pmatrix} 0.82 & -0.58 & 0 & 0 \\ 0.58 & 0.82 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$A^T \cdot A = \begin{pmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 8 & 20 \\ 5 & 13 \end{pmatrix}$$

$$p_a(\lambda) = \left\| \begin{matrix} 8-\lambda & 20 \\ 5 & 13-\lambda \end{matrix} \right\|$$

$$= (8-\lambda)(13-\lambda) - 100$$

$$= \lambda^2 - 21\lambda + 4$$

$$\lambda_1 = \frac{21 + 5\sqrt{17}}{2}, \quad \lambda_2 = \frac{21 - 5\sqrt{17}}{2}$$

$$\Rightarrow V = \begin{pmatrix} 0.4 & -0.91 \\ 0.91 & 0.4 \end{pmatrix}$$

The singular value decomposition is hence

$$\begin{pmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \underbrace{\begin{pmatrix} 0.82 & -0.58 & 0 & 0 \\ 0.58 & 0.82 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} 5.47 & 0 \\ 0 & 0.37 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}}_{\Sigma} \cdot \underbrace{\begin{pmatrix} 0.4 & 0.91 \\ -0.91 & 0.4 \end{pmatrix}}_{V^T}$$

1.3 Singular Value Decomposition - Theoretical

1. Show that the rank of A is r, where r is the minimum such that $\arg\max_{\substack{v \perp v_1, v_2, \dots, v_r \\ |v| = 1}} |Av| = 0$

Sketch of the proof:

By definition

the i-th singular vector is defined by

$$v_i = \arg\max_{\substack{v \perp v_1, v_2, \dots, v_{i-1}, \\ |v| = 1}} |A\underline{v}|$$

the construction process stops when

$$\arg\max_{\substack{v \perp v_1, v_2, \dots, v_r \\ |v| = 1}} |A\underline{v}| = 0.$$

we can then decompose A into a sum of rank-1 matrices

$$A = \sum_{i=1}^{r} \sigma_i \, \underline{u}_i \, \underline{v}_i^T$$

the space hence spans r dimensions and A is hence of rank r.

■

1.3

2. Show that $|\underline{u}_1^T A| = \max_{|u|=1} |\underline{u}^T A| = \sigma_1$

We begin by using the uniqueness of the SVD

$$A = U \Sigma V^T$$

$$\Rightarrow \|A\|_F^2 = tr(A A^T)$$

$$= tr(U \Sigma \Sigma^T U^T)$$

$$= \sum_i \sigma_i^2 \underbrace{u_i u_i^T}_{\substack{=1 \\ \text{due to ortho-} \\ \text{normality}}} = \sum_i \sigma_i^2$$

similarly we have

$$\|A\|_2^2 = \max_v \frac{\|Av\|}{\|v\|} = \max_{\|v\|=1} \|Av\|$$

$$\Rightarrow |A^T|_2^2 = \max_{\|u\|=1} \|A^T u\|$$

$$= \max_{\|u\|=1} \|V \Sigma^T U^T u\|$$

using the invariance under orthogonal transformations

$$= \max_{\|u\|=1} \|\Sigma^T U^T u\|$$

$$= \max_{\|v\|=1} \|\Sigma^T v\|$$

$$= \sigma_1$$

using the following definition combined with the orthonormality of the left-singular vectors then completes the proof

$$\|A v_1\| = \arg \max_{\|\underline{v}\|=1} \|A\underline{v}\| \Rightarrow \|u_1^T A\| = \arg \max_{\|u\|=1} \|\underline{u}^T A\|.$$

$\blacksquare$

2 Probability Theory

2.1 Variance of a Sum

Show that the variance of a sum is $var[x+y] = var[x] + var[y] + 2\,cov(x,y)$, where $cov(x,y)$ is the covariance between x and y.

$$\cancel{var(x+y) = \mathbb{E}\left((x+y)^2\right) - \left[\mathbb{E}(x+y)\right]^2}$$

Using $var(x+y) = cov(x+y, x+y)$

$$\Rightarrow var(X+Y) = cov(X+Y, X+Y)$$

$$= \mathbb{E}\left((X+Y)^2\right) - \mathbb{E}(X+Y)\mathbb{E}(X+Y)$$

$$= \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2 + \mathbb{E}(Y^2) - \left(\mathbb{E}(Y)\right)^2$$

$$+ 2\left(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right)$$

$$= var(X) + var(Y) + 2\,cov(X,Y).$$

$\blacksquare$

2.2 Pairwise Independence does not imply mutual Independence

We say that two random variables are pairwise independent if

$$p(X_2 \mid X_1) = p(X_2)$$

and hence

$$p(X_2, X_1) = p(X_1)\, p(X_2 \mid X_1) = p(X_1)\, p(X_2).$$

We say that n random variables are mutually independent if

$$p(X_i \mid X_S) = p(X_i) \qquad \forall\, S \subseteq \{1, \dots, n\} \setminus \{i\}$$

and hence

$$p(X_{1:n}) = \prod_{i=1}^{n} p(X_i).$$

Show that pairwise independence between all pairs of variables does not necessarily imply mutual independence. It suffices to give a counter example.

The standard counter-example to this comes from discrete set theory where one can construct the following counter-example on an event space of $\{1, 2, 3, 4\}$. Here we create 3 events, i.e.

$$A = \{1, 2\}, \quad B = \{1, 3\}, \quad C = \{1, 4\}$$

$$\Rightarrow \mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{1\}) = \frac{1}{4}$$

$$\mathbb{P}(A) \cdot \mathbb{P}(B)\, \mathbb{P}(C) = \frac{1}{8}$$

2.3 Bernoulli Distribution

The form of the Bernoulli distribution given by

$$\mathrm{Bern}(x \mid \mu) = \mu^x (1-\mu)^{1-x}$$

is not symmetric between the two values of x. In some situations, it will be more convenient to use an equivalent formulation for which $x \in \{-1, 1\}$, in which case the distribution can be written as

$$p(x \mid \mu) = \left(\frac{1-\mu}{2}\right)^{\frac{1-x}{2}} \left(\frac{1+\mu}{2}\right)^{\frac{1+x}{2}}$$

where $\mu \in [-1, 1]$. Show that the distribution is normalized, and evaluate its mean, variance, and entropy.

---

normalization $\sum_{x \in \{-1,1\}} p(x \mid \mu) = p(x = -1 \mid \mu) + p(x = 1 \mid \mu)$

$$= \left(\frac{1-\mu}{2}\right) + \left(\frac{1+\mu}{2}\right)$$

$$= \underline{\underline{1}}$$

the distribution is hence normalized.

mean: $E[x] = \sum_{x \in \{-1,1\}} x \, p(x \mid \mu)$

$$= (-1) \cdot \left(\frac{1-\mu}{2}\right) + \left(\frac{1+\mu}{2}\right)$$

$$= \underline{\underline{\mu}}$$

variance :

$$\mathrm{var}(x) = \mathbb{E}\left[(x-\mu)^2\right]$$

$$= \sum_{x \in \{-1, 1\}} (x-\mu)^2 \, p(x|\mu)$$

$$= (-1-\mu)^2 \left(\frac{1-\mu}{2}\right) + (1-\mu)^2 \left(\frac{1+\mu}{2}\right)$$

$$= \frac{1}{2}\left((1+\mu)^2 (1-\mu) + (1-\mu)^2 (1+\mu) \right)$$

$$= \frac{1}{2}\left(2 - 2\mu^2 \right)$$

$$= 1 - \mu^2$$

entropy :

$$H(x) = - \sum_{x \in \{-1, 1\}} p(x|\mu) \ln p(x|\mu)$$

$$= - \left(\frac{1-\mu}{2}\right) \ln \left(\frac{1-\mu}{2}\right) - \left(\frac{1+\mu}{2}\right) \ln \left(\frac{1+\mu}{2}\right)$$

$$= - \left(\frac{1-\mu}{2}\right)\left(\ln (1-\mu) - \ln(2)\right)$$

$$- \left(\frac{1+\mu}{2}\right)\left(\ln (1+\mu) - \ln(2)\right)$$

2.4 Beta Distribution

Prove that the beta distribution, given by

$$\mathrm{Beta}(\mu \mid a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1}$$

is correctly normalized, so that

$$\int_0^1 \mathrm{Beta}(\mu \mid a,b)\, d\mu = 1$$

holds. This is equivalent to showing that

$$\int_0^1 \mu^{a-1}(1-\mu)^{b-1}\, d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

~~From~~ Use expression 14 to prove 13 as follows. First bring the integral over y inside the integrand of the integral over x, next make the change of variable $t = y + x$ where x is fixed, then interchange the order of the x and t integrations, and finally make the change of variable $x = t\mu$ where t is fixed.

$$\Gamma(a)\Gamma(b) = \int_0^\infty \exp(-x)\, x^{a-1}\, dx \int_0^\infty \exp(-y)\, y^{b-1}\, dy$$

$$= \int_0^\infty \int_0^\infty \exp(-x-y)\, x^{a-1} y^{b-1}\, dy\, dx$$

variable change $t = y + x \Rightarrow dt = dy$

$$\cancel{= \int_0^\infty \int_x^\infty \exp(-t)\, x^{a-1} (t-x)^{b-1}\, dt\, dx}$$

$$= \int_0^\infty \exp(-t) \left(\int_0^t x^{a-1} (t-x)^{b-1}\, dx \right) dt$$

$$x = \mu t \Rightarrow dx = t \, d\mu$$

$$= \int_0^\infty \exp(-t) \left(t^{a+b-1} \int_0^1 \mu^{a-1} (1-\mu)^{b-1} \, d\mu \right) dt$$

$$= \int_0^\infty \exp(-t) \, t^{a+b-1} \, dt \cdot \int_0^1 \mu^{a-1} (1-\mu)^{b-1} \, d\mu$$

$$\Rightarrow \Gamma(a)\Gamma(b) = \int_0^\infty \exp(-t) \, t^{a+b-1} \, dt \cdot \int_0^1 \mu^{a-1} (1-\mu)^{b-1} \, d\mu$$

$$= \Gamma(a+b)$$

$$\Rightarrow \int_0^1 \mu^{a-1} (1-\mu)^{b-1} \, d\mu = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Rightarrow \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \mu^{a-1} (1-\mu)^{b-1} \, d\mu = 1$$

$\blacksquare$

2.6 Deriving the Inverse Gamma Density

Let $X \sim Ga(a,b)$, i.e.

$$Ga(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-xb}$$

Let $Y \sim \frac{1}{X}$. Show that $Y \sim IG(a,b)$, i.e.

$$IG(x \mid shape = a, scale = b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-b/x}$$

$\$$ Using a variable transformation for this 1-1 transformation we then have the Jacobian

$$\frac{dX}{dY} = -\frac{1}{Y^2}$$

$$\Rightarrow f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|$$

$$= \frac{b^a}{\Gamma(a)} \left(\frac{1}{y}\right)^{a-1} e^{-\frac{b}{y}} \left| -\frac{1}{y^2} \right|$$

$$= \frac{b^a}{\Gamma(a)} \left(\frac{1}{y}\right)^{a+1} e^{-b/y}$$

$$= \frac{b^a}{\Gamma(a)} y^{-(a+1)} e^{-b/y}$$

$\blacksquare$

2.7 Normalization Constant for a 1D Gaussian

The normalization constant for a zero-mean Gaussian is given by

$$Z = \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

where $a = -\infty$ and $b = \infty$. To compute this, consider its square

$$Z^2 = \int_a^b \int_a^b \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) dx\, dy$$

Let us change variables from cartesian (x, y) to polar (r, θ) using $x = r\cos\theta$ and $y = r\sin\theta$. Since $dx\,dy = r\,dr\,d\theta$, and $\cos^2(\theta) + \sin^2(\theta) = 1$, we have

$$Z^2 = \int_0^{2\pi} \int_0^{\infty} r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr\, d\theta$$

Evaluate this integral and hence show $Z = \sigma\sqrt{2\pi}$. (..)

$$Z^2 = \int_0^{2\pi} d\theta \cdot \int_0^{\infty} r \exp\left(-\frac{r^2}{2\sigma^2}\right) dr$$

$$= 2\pi \cdot \left[- \sigma^2 \exp\left(-\frac{r^2}{2\sigma^2}\right) \right]_{r=0}^{r=\infty}$$

$$= 2\pi \cdot \sigma^2 \ (0 + 1)$$

$$\Rightarrow \quad Z = \sigma\sqrt{2\pi}$$

$\blacksquare$

2.8 Kullback - Leibler Divergence

Evaluate the Kullback - Leibler divergence, expressing the relative entropy of two probability distributions,

$$KL(p \| q) = -\int p(x) \ln(q(x)) dx - \left(-\int p(x) \ln(p(x)) dx\right)$$

$$= -\int p(x) \ln\left\{\frac{q(x)}{p(x)}\right\} dx$$

between two Gaussians $p(x) = \mathcal{N}(x \mid \mu, \Sigma)$ and $q(x) = \mathcal{N}(x \mid m, L)$.

$$= \int \left(\ln(p(x)) - \ln(q(x))\right) p(x)\, dx$$

$$= \int \left[-\cancel{\frac{1}{2}\log(2\pi)} - \log(\Sigma) - \frac{1}{2}\left(\frac{x-\mu}{\Sigma}\right)^2 + \cancel{\frac{1}{2}\log(2\pi)} + \log(L) + \frac{1}{2}\left(\frac{x-m}{L}\right)^2\right] \cdot \frac{1}{\sqrt{2\pi}\cdot\Sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\Sigma}\right)^2\right) dx$$

$$= \int \left\{\log\left(\frac{L}{\Sigma}\right) + \frac{1}{2}\left[\left(\frac{x-m}{L}\right)^2 - \left(\frac{x-\mu}{\Sigma}\right)^2\right]\right\} \cdot \frac{1}{\sqrt{2\pi}\cdot\Sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\Sigma}\right)^2\right) dx$$

$$= \mathbb{E}_{p(x)}\left\{\log\left(\frac{L}{\Sigma}\right) + \frac{1}{2}\left[\left(\frac{x-m}{L}\right)^2 - \left(\frac{x-\mu}{\Sigma}\right)^2\right]\right\}$$

$$= \log\left(\frac{L}{\Sigma}\right) + \frac{1}{2L^2} \mathbb{E}_{p(x)}\left[(x-m)^2\right] - \frac{1}{2\Sigma^2} \mathbb{E}_{p(x)}\left[(x-\mu)^2\right]$$

$$= \log\left(\frac{L}{\Sigma}\right) + \frac{1}{2L^2} \mathbb{E}_{p(x)}\left[(x-m)^2\right] - \frac{1}{2}$$

we now complete the square in the expectation w.r.t. $p(x)$

$$= \log\left(\frac{L}{\Sigma}\right) + \frac{1}{2L^2}\,\mathbb{E}_{p(x)}\left[(X-\mu)^2 + 2(X-\mu)(\mu-m) + (\mu-m)^2\right]$$

$$- \frac{1}{2}$$

$$= \log\left(\frac{L}{\Sigma}\right) + \frac{1}{2L^2}\left[\mathbb{E}_{p(x)}\left[(x-\mu)^2\right] + 2(\mu-m)\,\mathbb{E}_{p(x)}\left[x-\mu\right] + (\mu-m)^2\right] - \frac{1}{2}$$

$$= \log\left(\frac{L}{\Sigma}\right) + \frac{\Sigma^2 + (\mu-m)^2}{2L^2} - \frac{1}{2}$$