Pekka Kasa

Homework 1

April 12, 2020

1 Chapter 1

1.1 Problem

Statement

Consider the sum-of-squares error function given by $E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$ in which the function $y(x, \mathbf{w})$ is given by the polynomial $P(x) = \sum_{j=0}^{M} w_j x^j$. Show that the coefficients $\mathbf{w} = \{w_i\}$ that minimize this error function are given by the solution to the following set of linear equations:

$$\sum_{j=0}^{M} A_{ij} w_j = T_i$$

where

$$A_{ij} = \sum_{n=1}^{N} (x_n)^{i+j} \qquad T_i = \sum_{n=1}^{N} (x_n)^i t_n$$

Solution

Let's start by reminding ourselves about what the different quantities in these equations stand for. This illustrates their dimensions at the same time.

- We have N training examples of the form (x_i, t_i) where $x_i, t_i \in \mathbb{R} \quad \forall i \in \{1, \dots, N\}$.
- We have $M + 1$ polynomial coefficients $w_j \in \mathbb{R} \quad \forall j \in \{0, \dots, M\}$
- The claim is that there exists N (the amount of training examples) linear equations whose solution is given by such coefficients in the equations that the function $E(\mathbf{w})$ gets its minimum value.
- The minimum for this specific (convex) function is attained when the gradient is zero. The function has only one local minimum, i.e. the global minimum.

$$\partial_{w_i} E(\mathbf{w}) = \sum_{j=0}^{M} A_{ij} w_j - T_i = \sum_{j=0}^{M}\sum_{n=1}^{N} (x_n)^{i+j} w_j - \sum_{n=1}^{N} (x_n)^i t_n = 0 \quad \forall i \in \{0, \dots, M\}$$

What we essentially want to do is to take the gradient $\nabla_{\mathbf{w}} E(\mathbf{w})$ with the appropriate form for the fitting function plugged in and show that the equation holds. We can get the gradient by taking the partials w.r.t. to each w_i separately and then building the gradient vector from those. Life can be made easier by calculating w_i for some $i \in \{0, \dots, M\}$ and then applying it to all i.

$$\frac{\partial}{\partial w_i}E(\mathbf{w}) = \frac{\partial}{\partial w_i}\frac{1}{2}\sum_{n=1}^{N}\left\{\left(\sum_{j=0}^{M} w_j x^j\right) - t_n\right\}^2$$

$$= \underbrace{\frac{1}{2}\sum_{n=1}^{N}\frac{\partial}{\partial w_i}}_{\text{Diff. is linear}}\left\{\left(\sum_{j=0}^{M} w_j x_n^j\right) - t_n\right\}^2 = \sum_{n=1}^{N}\left\{\left(\left(\sum_{j=0}^{M} w_j x_n^j\right) - t_n\right)\left(\frac{\partial}{\partial w_i}\sum_{j=0}^{M} w_j x_n^j - \frac{\partial}{\partial w_i}t_n\right)\right\}$$

$$= \sum_{n=1}^{N}\left\{\left(\left(\sum_{j=0}^{M} w_j x_n^j\right) - t_n\right)\frac{\partial}{\partial w_i}\sum_{j=0}^{M} w_j x_n^j\right\} = \sum_{n=1}^{N}\left\{\left(\left(\sum_{j=0}^{M} w_j x_n^j\right) - t_n\right)x_n^i\right\}$$

$$= \sum_{n=1}^{N}\left\{x_n^i\left(\left(\sum_{j=0}^{M} w_j x_n^j\right) - t_n\right)\right\} = \sum_{n=1}^{N}\left(\sum_{j=0}^{M} w_j x_n^{i+j}\right) - \sum_{n=1}^{N} x_n^i t_n$$

$$= \sum_{n=1}^{N}\sum_{j=0}^{M} x_n^{i+j} w_j - \sum_{n=1}^{N} x_n^i t_n = 0 \quad \square$$

1.2 Problem *

Statement

Derive the same set of equations for the ℓ^2 regularized case.

Solution

Take gradient of the error function $E : \mathbb{R}^D \to [0, \infty[$ where $y : \mathbb{R}^D \to \mathbb{R}$

$$\nabla_{\mathbf{w}} E(\mathbf{w}) = \nabla_{\mathbf{w}}\left\{\frac{1}{2}\sum_{n=1}^{N}(y(x_n,\mathbf{w}) - t_n)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2\right\} = \frac{1}{2}\sum_{n=1}^{N}\nabla_{\mathbf{w}}(y(x_n,\mathbf{w}) - t_n)^2 + \frac{\lambda}{2}\nabla_{\mathbf{w}}\|\mathbf{w}\|^2$$

$$= \lambda\mathbf{w} + \sum_{n=1}^{N}(y(x_n,\mathbf{w}) - t_n)\nabla_{\mathbf{w}} y(x_n,\mathbf{w})$$

1.3 Problem

Statement

Suppose that we have three colored boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes.

If a box is chosen at random with probabilities $p(r) = 0.2, p(b) = 0.2, p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

$$\text{Solve } P(X = a) \text{ and } p(Y = g|X = o)$$

Solution

$$P(Y = r) = 0.2, \quad P(Y = b) = 0.2, \quad P(Y = g) = 0.6$$
$$P(X = a|Y = r) = 0.3, \quad P(X = o|Y = r) = 0.4, \quad P(X = l|Y = r) = 0.3$$
$$P(X = a|Y = b) = 0.5, \quad p(X = o|Y = b) = 0.5, \quad P(X = l|Y = b) = 0.0$$
$$P(X = a|Y = g) = 0.3, \quad p(X = o|Y = g) = 0.3, \quad P(X = l|Y = g) = 0.4$$

$$p(X) = \sum_Y P(X, Y) = \sum_Y p(X|Y)p(Y)$$

$$\begin{aligned} p(X = a) &= \sum_{Y \in \{r,g,b\}} p(X = a|Y)p(Y) = P(a|r)p(r) + p(a|b)p(b) + p(a|g)p(g) \\ &= 0.3 \cdot 0.2 + 0.5 \cdot 0.2 + 0.3 \cdot 0.6 = 0.34 \end{aligned}$$

$$\begin{aligned} P(Y = g|X = o) &= \frac{P(X = o|Y = g)P(Y = g)}{P(X = o)} = \frac{P(X = o|Y = g)P(Y = g)}{\sum_{Y \in \{r,b,g\}} P(X = o|Y)P(Y)} \\ &= \frac{P(X = o|Y = g)P(Y = g)}{P(X = o|Y = r)P(Y = r) + P(X = o|Y = b)P(Y = b) + P(X = o|Y = g)P(Y = g)} \\ &= \frac{0.3 \cdot 0.6}{0.4 \cdot 0.2 + 0.5 \cdot 0.2 + 0.3 \cdot 0.6} = \frac{0.18}{0.36} = 0.5 \end{aligned}$$

1.4 Problem *

Statement

Given a probability density $p_x(x)$ and a change of a variable $x = g(t)$, the new probability density $p_t(t)$ will have its mode at a different place than the original, in general.

The change of variables for probability densities includes the Jacobian determinant:

$$p_t(t) = p_x(x) \left| \frac{d}{dt} x \right| = p_x(g(t)) \left| \frac{d}{dt} g(t) \right| = p_x(g(t)) \left| g'(t) \right|$$

The claim is that you can't get the maximum for the density p_t from the known place of the maximum for p_X by simply doing the inverse variable change:

$$t_* = g^{-1}(x_*)$$

Solution

The solution is to take the derivative of the new density $p_t'(t_*) = 0$ and figure out where its zeros lie - these correspond to the maximum of the new probability density function. We need to compare them to the zeros of the original probability density function $p_x'(x_*) = 0$.

1.5 Problem

Statement

Using the definition

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

show that var$[f]$ satisfies

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

Solution

$$f(x) = x$$

$$\begin{aligned}
\mathrm{var}[x] &= \mathbb{E}\left[(x - \mathbb{E}\left[x\right])^2 \right] = \mathbb{E}\left[x^2 - 2x\mathbb{E}\left[x\right] + \mathbb{E}\left[x\right]^2 \right] \\
&= \mathbb{E}\left[x^2\right] - 2\mathbb{E}\left[x \underbrace{\mathbb{E}\left[x\right]}_{\in \mathbb{R}} \right] + \mathbb{E}\left[\mathbb{E}\left[x\right]^2 \right] \\
&= \mathbb{E}\left[x^2\right] - 2\mathbb{E}\left[x\right] \mathbb{E}\left[x\right] + \mathbb{E}\left[x\right]^2 \\
&= \mathbb{E}\left[x^2\right] - 2\mathbb{E}\left[x\right]^2 + \mathbb{E}\left[x\right]^2 = \mathbb{E}\left[x^2\right] - \mathbb{E}\left[x\right]^2 \qquad \square
\end{aligned}$$

1.6 Problem

Statement

Show that if two variables x and y are independent, then their covariance is zero.
Given

$$p(x, y) = p(x)p(y)$$

Prove

$$\text{cov}[x, y] = \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] = 0$$

Solution

The same precise logic applies for both the discrete and continuous case where we simply swap the integral sign to a sum.

$$\begin{aligned}\mathbb{E}_{x,y}[xy] &= \int_{x,y} xy \cdot p(x, y)\mathrm{d}V = \int_X \int_Y xy \cdot p(x, y)\mathrm{d}y\mathrm{d}x = \int_X \int_Y xy \cdot p(x)p(y)\mathrm{d}y\mathrm{d}x \\ &= \int_X xp(x)\mathrm{d}x \int_Y yp(y)\mathrm{d}y = \mathbb{E}[x]\mathbb{E}[y] \qquad \square\end{aligned}$$

1.7 Problem

Statement

In this exercise, we prove the normalization condition $\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)dx = 1$ for the univariate Gaussian.
To do this, consider the integral

$$I = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right)\mathrm{d}x$$

which we can evaluate by first writing its square in the form

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}y^2\right)\mathrm{d}x\mathrm{d}y$$

Now make the transformation from Cartesian coordinates (x, y) to polar coordinates (r, θ) and then substitute $u = r^2$. Show that, by performing the integrals over θ and u, and then taking the square root of both sides, we obtain

$$I = (2\pi\sigma^2)^{1/2}$$

Finally, use this result to show that the Gaussian distribution $\mathcal{N}(x|\mu, \sigma^2)$ is normalized.

Solution

$$\begin{cases} r = & \sqrt{x^2+y^2} \\ \theta = & \arctan x/y \end{cases} \Rightarrow \begin{cases} r^2 = & x^2+y^2 \\ \theta = & \arctan x/y \end{cases}$$

Let's rewrite the following in polar coordinates. This means rewriting

$$\begin{aligned}
I^2 &= \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}r^2\right) r\mathrm{d}r = \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}r^2\right) r\mathrm{d}r \\
&= \int_0^{2\pi} \mathrm{d}\theta \cdot -\sigma^2 \int_0^{\infty} -\frac{2r}{2\sigma^2} \cdot \exp\left(-\frac{1}{2\sigma^2}r^2\right) \mathrm{d}r = [2\pi - 0]\left[-\sigma^2 \cdot \exp\left(-\frac{1}{2\sigma^2}r^2\right)\Big|_0^{\infty}\right] \\
&= -2\pi\,\sigma^2 \cdot \exp\left(-\frac{1}{2\sigma^2}r^2\right)\Big|_0^{\infty} = -2\pi\sigma^2 \left[\lim_{r\to\infty} \exp\left(-\frac{1}{2\sigma^2}r^2\right) - \lim_{r\to 0} \exp\left(-\frac{1}{2\sigma^2}r^2\right)\right] \\
&= -2\pi\sigma^2 \left[\lim_{r\to\infty} \frac{1}{e^{\frac{r^2}{2\sigma^2}}} - \lim_{r\to 0} \frac{1}{e^{\frac{r^2}{2\sigma^2}}}\right] = -2\pi\sigma^2\,[0-1] = 2\pi\sigma^2
\end{aligned}$$

$$I = \sqrt{2\pi\sigma^2}$$

The result for the area of the Gaussian was only derived in the case where $\mu = 0$. We need to now show that the area of the integral is independent of μ.

$$t = x - \mu$$

$$\frac{d}{dx}t(x) = \frac{d}{dx}(x-\mu) = 1 \Rightarrow dt = dx \Rightarrow \int_{-\infty}^{\infty} \mathcal{N}(x|0,\sigma^2)dx = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)dx = 1 \quad \square$$

1.8 Problem

Statement

By using a change of variables, verify that the univariate Gaussian distribution given by

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

satisfies

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)x dx = \mu$$

Next, by differentiating both sides of the normalizing condition

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)dx = 1$$

with respect to σ^2, verify that the Gaussian satisfies

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)x^2 dx = \mu^2 + \sigma^2$$

Finally, show that

$$var[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

Solution

$$\begin{aligned}
\mathbb{E}[x] &= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} xdx = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}t^2\right\} (t+\mu)dt \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}t^2\right\} tdt + \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}t^2\right\} dt}_{=1} = I + \mu
\end{aligned}$$

$$\begin{aligned}
I &= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}t^2\right\} tdt = \frac{1}{(2\pi\sigma^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}t^2\right\} tdt \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}t^2\right\} tdt = \frac{2}{(2\pi\sigma^2)^{1/2}} \left[\underbrace{\lim_{t\to\infty} \frac{1}{e^{\frac{t^2}{2\sigma^2}}}}_{=-1} - \underbrace{\lim_{t\to-\infty} \frac{1}{e^{\frac{t^2}{2\sigma^2}}}}_{=-1} \right] = 0
\end{aligned}$$

$$\therefore \mathbb{E}[x] = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} xdx = I + \mu = \mu \quad \Box_1$$

$$
\begin{aligned}
\frac{d\cdot 1}{d\sigma^2}\cdot &= \frac{d}{d\sigma^2}\int_{-\infty}^{\infty}\mathcal{N}(x|\mu,\sigma^2)dx = \frac{d}{d\sigma^2}\int_{-\infty}^{\infty}\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\frac{d}{d\sigma^2}\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\left[\frac{d}{d\sigma^2}\frac{1}{(2\pi\sigma^2)^{1/2}}\right]\cdot\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} + \frac{1}{(2\pi\sigma^2)^{1/2}}\cdot\frac{d}{d\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\left[-\frac{1}{\sqrt{8\pi\sigma^3}}\right]\cdot\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} + \frac{1}{(2\pi\sigma^2)^{1/2}}\cdot\frac{(x-\mu)^2}{2\sigma^4}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\mathrm{d}x \\
&= -\frac{1}{2\sigma^2}\underbrace{\int_{-\infty}^{\infty}\mathcal{N}(x|\mu,\sigma^2)\mathrm{d}x}_{=1} + \int_{-\infty}^{\infty}\frac{x^2-2x\mu+\mu^2}{2\sigma^4}\underbrace{\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}}_{=\mathcal{N}(x|\mu,\sigma^2)}\mathrm{d}x \\
&= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\left[\underbrace{\int_{-\infty}^{\infty}x^2\mathcal{N}(x|\mu,\sigma^2)\mathrm{d}x}_{=\mathbb{E}[x^2]} - 2\mu\underbrace{\int_{-\infty}^{\infty}x\mathcal{N}(x|\mu,\sigma^2)\mathrm{d}x}_{=\mu} + \mu^2\underbrace{\int_{-\infty}^{\infty}\mathcal{N}(x|\mu,\sigma^2)\mathrm{d}x}_{=1}\right] \\
&= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\left[\mathbb{E}[x^2]-\mu^2\right] \iff \frac{2\sigma^4}{2\sigma^2}+\mu^2 = \mathbb{E}[x^2] \iff \mathbb{E}[x^2] = \mu^2+\sigma^2 \quad \square
\end{aligned}
$$

1.9 Problem

Statement

Show that the mode (i.e. the maximum) of the Gaussian distribution

$$
\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}
$$

is given by μ.
Similarly, show that the mode of the multivariate Gaussian

$$
\mathcal{N}(\mathbf{x}|\mu,\mathbf{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right\}
$$

is given by $\mu \in \mathbb{R}^D$.

Solution

$$
\begin{aligned}
\frac{d}{dx}\mathcal{N}(x|\mu,\sigma^2) = \frac{d}{dx}\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} = \frac{2(x-\mu)}{(2\pi\sigma^2)^{1/2}}\underbrace{\mathcal{N}(x|\mu,\sigma^2)}_{>0} \\
= 2(x-\mu) = 0 \iff x = \mu \quad \square_1
\end{aligned}
$$

$$\nabla_x \mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma}) = \nabla_x \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right\} = -\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\underbrace{\mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma})}_{>0\ \forall x_i \in \mathbf{x}}$$
$$= \mathbf{x} - \mu = 0 \in \mathbb{R}^D \iff x = \mu \quad \square_2$$

1.10 Problem

Statement

Suppose that the two variables x and z are statistically independent. Show that the mean and variance of their sum satisfies

$$\mathbb{E}[x+z] = \mathbb{E}[x] + \mathbb{E}[z]$$
$$\mathrm{var}[x+z] = \mathrm{var}[x] + \mathrm{var}[z]$$

Solution

This is by definition. The expectations of both independent and dependent variables is linear. To give a detailed proof in integral form of the independent case:

$$\begin{aligned}
\int_V p(x,z)[x+z]\mathrm{d}V &= \int_V p(x)xp(z) + p(z)zp(x)\mathrm{d}V = \int_x \int_z p(x)xp(z) + p(z)zp(x)\mathrm{d}z\mathrm{d}x \\
&= \int_x \left[p(x)x \int_z p(z)\mathrm{d}z + p(x)\int_z p(z)z\mathrm{d}z\right]\mathrm{d}x \\
&= \int_x \left[p(x)x \int_z p(z)\mathrm{d}z + p(x)\mathbb{E}[z]\right]\mathrm{d}x \\
&= \int_x p(x)\left[x \int_z p(z)\mathrm{d}z + \mathbb{E}[z]\right]\mathrm{d}x \\
&= \int_x p(x)\left[x + \mathbb{E}[z]\right]\mathrm{d}x = \int_x p(x)x\mathrm{d}x + \mathbb{E}[z]\int_x p(x)\mathrm{d}x = \mathbb{E}[x] + \mathbb{E}[z]
\end{aligned}$$

$$
\begin{aligned}
\operatorname{var}[x+z] &= \mathbb{E}[(x+z)^2] - \mathbb{E}[x+z]^2 = \mathbb{E}[x^2+2xz+z^2] - (\mathbb{E}[x]+\mathbb{E}[z])^2 \\
&= \mathbb{E}[x^2] + 2\mathbb{E}[xz] + \mathbb{E}[z^2] - \mathbb{E}[x]^2 - 2\mathbb{E}[x]\mathbb{E}[z] - \mathbb{E}[z]^2 \\
&= \mathbb{E}[x^2] - \mathbb{E}[x]^2 + \mathbb{E}[z^2] - \mathbb{E}[z]^2 + 2\underbrace{(\mathbb{E}[xz] - \mathbb{E}[x]\mathbb{E}[z])}_{=0 \text{ because of ind.}} \\
&= \operatorname{var}[x] + \operatorname{var}[z] \quad \square
\end{aligned}
$$

1.11 Problem

Statement

By setting the derivatives of the log likelihood function

$$
\ln p(\mathbf{x}|\mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln 2\pi
$$

with respect to μ and σ^2 to zero, verify the results

$$
\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n
$$

and

$$
\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2
$$

Solution

$$
\begin{aligned}
\partial_\mu \ln p(\mathbf{x}|\mu,\sigma^2) &= \partial_\mu \left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln 2\pi\right) = \sum_{n=1}^{N}\partial_\mu(x_n-\mu)^2 \\
&= \sum_{n=1}^{N}(x_n-\mu) = \sum_{n=1}^{N}x_n - \sum_{n=1}^{N}\mu = \sum_{n=1}^{N}x_n - N\mu = 0 \iff \mu = \frac{1}{N}\sum_{n=1}^{N}x_n \quad \square
\end{aligned}
$$

$$\begin{aligned}
\partial_{\sigma^2} \ln p(\mathbf{x}|\mu, \sigma^2) &= \partial_{\sigma^2} \left(-\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln 2\pi \right) \\
&= -\frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^2 \cdot \partial_{\sigma^2} \frac{1}{\sigma^2} - \frac{N}{2} \cdot \partial_{\sigma^2} \ln \sigma^2 \\
&= \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^2 \frac{1}{\sigma^4} - \frac{N}{2\sigma^2} = 0 \iff \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2
\end{aligned}$$

$$\therefore \sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{\text{ML}})^2$$

1.12 Problem

Statement

Using the results

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x \mathrm{d}x = \mu$$

and

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x^2 \mathrm{d}x = \mu^2 + \sigma^2$$

show that

$$\mathbb{E}[x_n x_m] = \mu^2 + I_{nm} \sigma^2$$

where x_n and x_m denote data points sampled from a Gaussian distribution with mean μ and variance σ^2, and I_{nm} satisfies $I_{nm} = 1$ if $n = m$ and $I_{nm} = 0$ otherwise.
Hence prove

$$\mathbb{E}[\mu_{\text{ML}}] = \mu$$

and

$$\mathbb{E}[\sigma^2_{\text{ML}}] = \left(\frac{N-1}{N} \right) \sigma^2$$

Solution