<Machine Learning HW #1>

21300008 강민수

1.5. Using the definition (1.38) show that $\mathrm{var}[\mathbf{f}(\mathrm{x})]$ satisfies (1.39).

Definition (1.38.): $\mathrm{var}[f] = \mathbb{E}[\,(f(x) - \mathbb{E}[f(x)])^2\,]$

Definition (1.39.): $\mathrm{var}[f] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$

Proof)

X가 discrete random variable인지, continuous random variable인지에 따라, expectation의 정의가 달라지고, 그 증명 과정이 달라진다. 따라서 x가 discrete인 경우와 continuous인 경우로 나누어 증명을 수행하였다.

(1) x가 discrete일 때

x가 discrete random variable인 경우 any real valued probability mass function f(x)의 기대값, $\mathbb{E}[f(x)]$는 $\sum_x f(x)p(x)$로 나타낼 수 있다. (이는 증명 대상이 아니므로 생략한다.) 이를 μ라 하자. 그렇다면 definition 1.38은 아래와 같이 전개할 수 있다.

$$\mathbb{E}[(f(x)-\mu)^2] = \textstyle\sum_x\{(f(x)-\mu)^2 p(x)\} = \sum_x\{f(x)^2 p(x) - 2\mu f(x)p(x) + \mu^2 p(x)\}$$
$$= \textstyle\sum_x\{f(x)^2 p(x)\} - \sum_x\{2\mu f(x)p(x)\} + \sum_x\{\mu^2 p(x)\}$$
$$= \textstyle\sum_x\{f(x)^2 p(x)\} - 2\mu\sum_x\{f(x)p(x)\} + \mu^2\sum_x\{p(x)\}$$

그런데, $\sum_x\{p(x)\}$는 "random variable x에 대한 표본공간의 모든 확률의 합은 1"이라는 확률 공리에 의해 1이 된다. 따라서, 확률 공리와 expectation의 정의에 의해 아래와 같은 식이 도출된다.

$$= \mathbb{E}[f(x)^2] - 2\mu\mathbb{E}[f(x)] + \mu^2 \qquad (\mathbb{E}[f(x)]\text{를 }\mu\text{라고 두었으므로})$$
$$= \mathbb{E}[f(x)^2] - 2\mu^2 + \mu^2$$
$$= \mathbb{E}[f(x)^2] - \mu^2$$
$$= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 \qquad (\text{증명 끝})$$

따라서, x가 discrete일 때, definition 1.38으로 definition 1.39를 도출할 수 있다.

(2) x가 continuous일 때,

x가 continuous random variable일때, any real valued probability density function f(x)의 기대값, $\mathbb{E}[\mathrm{f(x)}]$는 $\int \mathrm{f(x)p(x)}dx$로 나타낼 수 있다. (이는 증명 대상이 아니므로 생략 한다.) 이를 μ라 하자. 그렇다면 definition 1.38은 아래와 같이 전개할 수 있다.

$$\begin{aligned}\mathbb{E}[(\mathrm{f(x)}-\mu)^2] &= \int (f(x)-\mu)^2 p(x)dx = \int \{f(x)^2 - 2\mu f(x) + \mu^2\} p(x)dx \\ &= \int f(x)^2 p(x) - 2\mu f(x)p(x) + \ \mu^2 p(x)\ dx \\ &= \int f(x)^2 p(x)\ dx - 2\mu \int f(x)p(x)dx + \mu^2 \int p(x)dx\end{aligned}$$

그런데, $\int p(x)dx$는 "random variable x에 대한 표본공간의 모든 확률의 합은 1"이라는 확률 공리에 의해 1이 된다. 따라서, 확률 공리와 expectation의 정의에 의해 아래와 같은 식이 도출된다.

$$= \int f(x)^2 p(x)\ dx - 2\mu \int f(x)p(x)dx + \mu^2 \qquad (\mathbb{E}[\mathrm{f(x)}]\text{를 }\mu\text{라고 두었으므로})$$

$$= \int f(x)^2 p(x)\ dx - 2\mu^2 + \mu^2$$

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

$$= \mathbb{E}[f(x)^2] - \mathbb{E}[\mathrm{f(x)}]^2 \qquad \text{(증명 끝)}$$

따라서, x가 continuous일 때, definition 1.38으로 definition 1.39를 도출할 수 있다.

(3) 결론

따라서, (1)과 (2)에 의해 x가 continuous random variable이거나 discrete random variable 일 때, $\mathrm{var[f]} = \mathbb{E}[\ (\mathrm{f(x)} -\ \mathbb{E}[\mathrm{f(x)}])^2\ ] =\ \mathbb{E}[f(x)^2] - \mathbb{E}[\mathrm{f(x)}]^2$이 성립함을 증명할 수 있다.

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

Proof)

x와 y의 covariance는 아래 공식으로 구할 수 있다.

$$\mathrm{covariance[x,y]} =\ \mathbb{E}_{\mathbf{x,y}}[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\boldsymbol{y} - \mathbb{E}[\mathbf{y}]\}] =\ \mathbb{E}_{\mathbf{x,y}}[\mathbf{xy}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]$$

x와 y가 독립이므로, xy의 기댓값은 x와 y 각각의 기댓값의 곱, 즉 $\mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}]$와 같다. 따라서, 아래와 같은 결과가 도출된다.

$$\mathrm{covariance[x,y]} =\ \mathbb{E}_{\mathbf{x,y}}[\mathbf{xy}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}] =\ \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}] = 0$$

그러므로, x와 y가 서로 독립일 때, x와 y의 covariance는 0이다. (증명 끝)

1.9. Show that the mode (i. e. the maximum) of the Gaussian distribution (1.46) is given by μ. Similarly, show that the mode of the multivariate Gaussian (1.52) is given by μ.

Proof)

(1) Proof for univariate Gaussian distribution (1. 46.)

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \times \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\} \qquad (1.\ 46.)$$

Proof)

위의 Gaussian distribution의 식(1. 46.)을 미분하면 아래와 같다.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dx}}\{N(x \mid \mu, \sigma^2)\} &= \frac{d}{dx}\left[\frac{1}{\sigma\sqrt{2\pi}} \times \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}\right] \\ &= \frac{1}{\sigma\sqrt{2\pi}} \times \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \times \left(-\frac{1}{2\sigma^2} \times 2(x-\mu)\right) \\ &= -\frac{1}{\sigma^3\sqrt{2\pi}}(x-\mu) \times \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\end{aligned}$$

따라서, $x = \mu$에서 미분 값이 0이 되고 $\mathrm{x} > \mu, x = \mu, x < \mu$ 각각의 경우에 따라 나오는 $N(x \mid \mu, \sigma^2)$의 값에 따라 $\mathrm{x} = \mu$에서 최댓값을 갖는지 증명할 수 있다. 이를 증명하면 아래와 같다.

$\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$를 $f(x)$라고 하자. 표준편차 σ와 원주율 상수 π의 정의에 의해, $\sigma > 0, \pi > 0$이므로, 아래 ①, ②, ③ 경우로 나누어 판단해야 한다.

① $x > \mu$일 때,

$$0 < f(x) < 1$$

② $x < \mu$일 때,

$$0 < f(x) < 1$$

③ $x = \mu$일 때,

$$f(x) = 1$$

따라서, $x = \mu$일 때 Gaussian distribution (1. 46.)의 mode를 얻을 수 있다.

(증명끝)

(2) Proof of multivariate Gaussian (1. 52)

$$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(|\Sigma|)^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\} \qquad (1.\ 52.)$$

Proof)

$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(|\Sigma|)^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$의 양변을 미분하면 아래와 같

다.

$$\frac{d}{dx}\{N(x \mid \mu, \Sigma)\} = \frac{d}{dx}\left[\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(|\Sigma|)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}\right]$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{(|\Sigma|)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} \times \frac{d}{dx}\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

$f(x) = x^TAx$일 때, $\frac{\mathrm{d}f(x)}{dx} = (A^T + A)x$이므로,

$$\frac{d}{dx}\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] = -\frac{1}{2}\{(\Sigma^{-1})^T + \Sigma^{-1}\}(x-\mu)$$

그런데, $(\Sigma^{-1})^T = \Sigma^{-1}$이므로,

$$-\frac{1}{2}\{(\Sigma^{-1})^T + \Sigma^{-1}\}(x-\mu) = -\Sigma^{-1}(x-\mu)$$

$$\therefore \frac{d}{dx}\{N(x \mid \mu, \Sigma)\} = N(x \mid \mu, \Sigma) \times \{-\Sigma^{-1}(x-\mu)\}$$

따라서, $x = \mu$에서 미분 값이 0이 되고 최댓값을 갖기 때문에 $x = \mu$에서 Gaussian distribution (1. 46.)의 mode를 얻을 수 있다.

(증명 끝)

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

$$\mathbb{E}[\mathbf{x}+\mathbf{z}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{z}] \qquad \textbf{(1. 128.)}$$

$$\mathbf{Var}[\mathbf{x}+\mathbf{z}] = \mathbf{Var}[\mathbf{x}] + \mathbf{Var}[\mathbf{z}] \qquad \textbf{(1. 129.)}$$

Proof)

x와 z가 independent이므로, $p_{x,z}(xz) = p(x)p(z)$이다. 이 사실을 이용하여 1.128.과 1.129.를 증명하면 다음과 같다.

(1) x, z가 discrete일 경우의 $\mathbb{E}[\mathbf{x}+\mathbf{z}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{z}]$ 증명

$$\mathbb{E}[\mathbf{x}+\mathbf{z}] = \sum_x\sum_z(x+z)p_{x,z}(x,z) = \sum_x\sum_z(x+z)p(x)p(z)$$
$$= \sum_x\sum_z xp(x)p(z) + \sum_x\sum_z zp(x)p(z)$$

확률공리에 의해, $\sum_x\sum_z xp(x)p(z)$ 에서 $\sum_z p(z) = 1$ 이고, $\sum_x\sum_z zp(x)p(z)$ 에서 $\sum_x p(x) = 1$이므로

$$\sum_x\sum_z xp(x)p(z) + \sum_x\sum_z zp(x)p(z) = \sum_x xp(x) + \sum_z zp(z) = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{z}]$$

$$\therefore \mathbb{E}[\mathbf{x}+\mathbf{z}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{z}]$$

(증명 끝)

(2) x, z가 continuous일 경우의 $\mathbb{E}[\mathbf{x}+\mathbf{z}] = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{z}]$ 증명

$$\begin{aligned}\mathbb{E}[\mathbf{x}+\mathbf{z}] &= \int\int (x+z)p_{x,z}(x,z)dx\,dz = \int\int (x+z)p_x(x)p_z(z)dx\,dz \\ &= \int\int xp_x(x)p(z)dz\,dx + \int\int zp_x(x)p_z(z)dx\,dz \\ &= \int xp_x(x)\,dx + \int zp_z(z)\,dz = \mathbb{E}[\mathbf{x}] + \mathbb{E}[\mathbf{z}]\end{aligned}$$

(증명 끝)

(3) $\mathbf{Var}[\mathbf{x}+\mathbf{z}] = \mathbf{Var}[\mathbf{x}] + \mathbf{Var}[\mathbf{z}]$의 증명

Var[x+z]는 아래와 같이 정의될 수 있다.

$$\mathrm{Var}[\mathrm{x}+\mathrm{z}] = \mathrm{Var}[\mathrm{x}] + Var[z] + 2Cov[x,z]$$

그런데, x, z가 independent이므로 Cov[x, z] = 0이다. 따라서, 결론은 아래와 같다.

$$\therefore \mathrm{Var}[\mathrm{x}+\mathrm{z}] = \mathrm{Var}[\mathrm{x}] + Var[z]$$

(증명 끝)

1.11. By setting the derivatives of the log likelihood function (1.54) with respect to μ and σ^2 equal to zero, verify the results (1.55) and (1.56)

$$\ln p\left(\mathbf{x}|\mu,\sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi). \qquad (1.54)$$

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}x_n \qquad (1.55)$$

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2 \qquad (1.56)$$

proof of 1.55)

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

$$\sum_{n=1}^{N}(\mu-x_n) = 0 \quad\Rightarrow\quad N\mu + \sum_{n=1}^{N}-x_n = 0 \quad\Rightarrow\quad \mu = \frac{1}{N}\sum_{n=1}^{N}x_n$$

$$\therefore \boldsymbol{\mu_{ML}} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} \boldsymbol{x_n}$$

(증명 끝)

proof of 1.56)

$$\frac{d}{d\sigma^2}\{ln\, p(x \mid \mu_{ML}, \sigma^2)\} = \frac{d}{d\sigma^2}\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu_{ML})^2 - \frac{N}{2}ln\,\sigma^2 - \frac{N}{2}\ln(2\pi)\right\}$$

$$= -\frac{1}{2}\sum_{n=1}^{N}(x_n - \mu_{ML})^2 \frac{d}{d\sigma^2}\left\{\frac{1}{\sigma^2}\right\} - \frac{N}{2}\frac{d}{d\sigma^2}\{ln\,\sigma^2\}$$

$$= -\frac{1}{2}\sum_{n=1}^{N}(x_n - \mu_{ML})^2\left\{-\frac{1}{\sigma^4}\right\} - \frac{N}{2\sigma^2} = 0$$

$$\frac{N}{2\sigma^2} = \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n - \mu_{ML})^2 \quad \Rightarrow \quad N\sigma_{ML}^2 = \sum_{n=1}^{N}(x_n - \mu_{ML})^2 \quad \Rightarrow \quad \sigma_{ML}^2 = \frac{1}{\mathrm{N}}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$$

$$\therefore \boldsymbol{\sigma_{ML}^2} = \frac{\mathbf{1}}{\mathbf{N}}\sum_{\boldsymbol{n=1}}^{\boldsymbol{N}}(\boldsymbol{x_n} - \boldsymbol{\mu_{ML}})^2$$

(증명 끝)

1.13. Suppose that the variance of a Gaussian is estimated using the result (1.56) but with the maximum likelihood estimate $\boldsymbol{\mu_{ML}}$ replaced with the true value $\boldsymbol{\mu}$ of the mean. Show that this estimator has the property that its expectation is given by the true variance $\boldsymbol{\sigma^2}$.

$$\sigma_{\mathrm{ML}}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2 \qquad (1.56)$$

Proof)

$\sigma_{ML}^2 = \frac{1}{\mathrm{N}}\sum_{n=1}^{N}(x_n - \mu_{ML})^2$에서 μ_{ML}을 μ로 대체한 식은 아래와 같다.

$$\frac{1}{\mathrm{N}}\sum_{n=1}^{N}(x_n - \mu)^2$$

그리고 이는 곧 σ^2를 구하는 식과 같다. 따라서 아래의 결론을 도출할 수 있다.

$$\therefore \boldsymbol{\sigma^2} = \frac{\mathbf{1}}{\mathbf{N}}\sum_{\boldsymbol{n=1}}^{\boldsymbol{N}}(\boldsymbol{x_n} - \boldsymbol{\mu})^2$$

(증명 끝)