Machine Learning HW 1

21300126 Seungwoo Kim

MNIST mean, variance and eigenvector graph

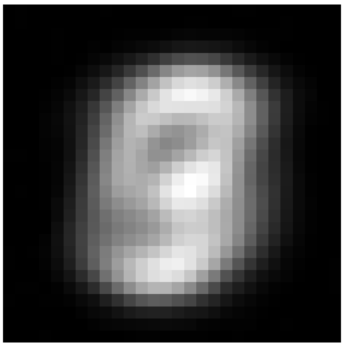

Figure 1. Mean of train_x

Figure 2. Variance of train_x

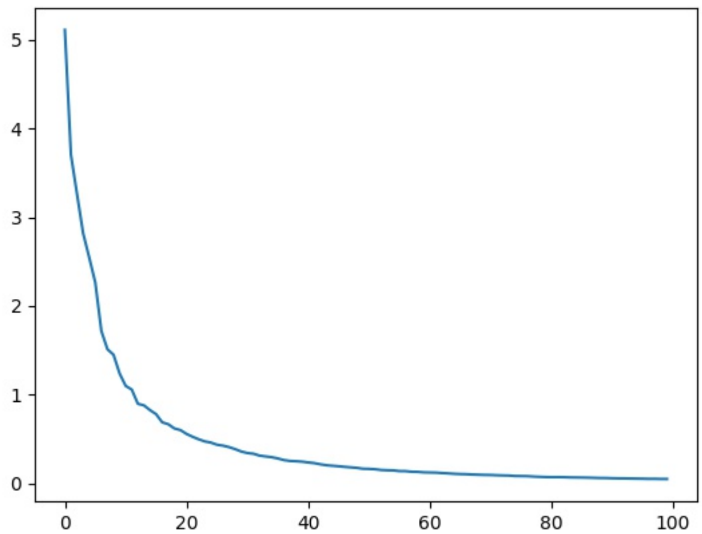

Figure 3. First 100 eigenvalues

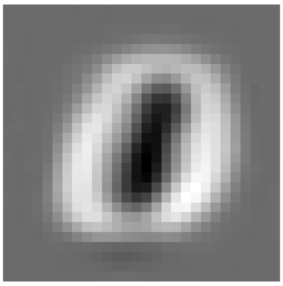

Figure 4. Eigenvector 1

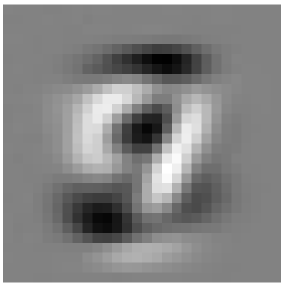

Figure 5. Eigenvector 2

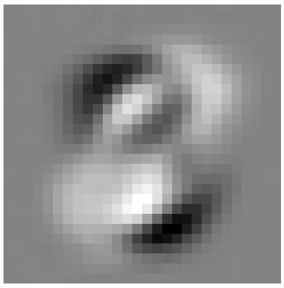

Figure 6. Eigenvector3

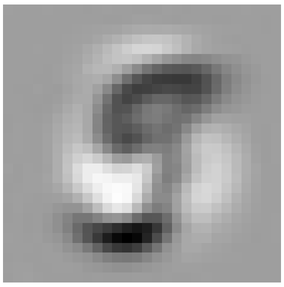

Figure 7. Eigenvector 4

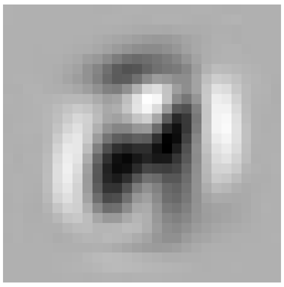

Figure 8. Eigenvector 5

Figure 9. Eigenvector 6

Figure 10. Eigenvector 7

Figure 11. Eigenvector 8

Figure 12. Eigenvector 9

Figure 13. Eigenvector 10

Bishop

1.5 Using the definition (1.38) show that $var[f(x)]$ satisfies (1.39).

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

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

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

$$Cov(x,y) = \mathrm{E}[(x - \mathrm{E}[x])(y - \mathrm{E}[y])] = \mathrm{E}(xy - x\mathrm{E}[y] - y\mathrm{E}[x] + \mathrm{E}[x]\mathrm{E}[y])$$

$$= \mathrm{E}[xy] - 2\mathrm{E}[x]\mathrm{E}[y] + \mathrm{E}[x]\mathrm{E}[y] = \mathrm{E}[xy] - \mathrm{E}[x]\mathrm{E}[y]$$

if two variables are independent, then $E[xy] = E[x]E[y]$. Therefore,

$$Cov(x,y) = \mathrm{E}[xy] - \mathrm{E}[x]\mathrm{E}[y] = \mathrm{E}[x]\mathrm{E}[y] - \mathrm{E}[x]\mathrm{E}[y] = 0$$

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 multivariate Gaussian (1.52) is given by $\boldsymbol{\mu}$.

In order to calculate the mode, or the maximum value, of the Gaussian distribution, the simplest way is to take a derivative of it and set it to 0. It is well-known that the Gaussian distribution is bell-shaped, so that it is not necessary to check whether the second derivative is less than 0. Without concerning the constants, the derivative of single Gaussian function is

$$\frac{d}{dx}\mathcal{N}(x|\mu,\sigma^2) = -\mathcal{N}(x|\mu,\sigma^2)\frac{(x-\mu)}{2\sigma^2} = 0$$

For the derivative to be equal to 0, $x = \mu$ or $x \to \infty$. Therefore, the mode (the maximum) of the Gaussian distribution is equal to μ.

Similary, for the multivariate Gaussian distribution,

$$\frac{d}{d\mathrm{x}}\mathcal{N}(\mathrm{x}|\boldsymbol{\mu},\Sigma) = -\frac{1}{2}\mathcal{N}(\mathrm{x}|\boldsymbol{\mu},\Sigma)\nabla_{\mathrm{x}}\{(\mathrm{x}-\boldsymbol{\mu})^{\mathbf{T}}\Sigma^{-1}(\mathrm{x}-\boldsymbol{\mu})\} = \mathbf{0}$$

Since Σ^{-1} is symmetry, multiplying both side with Σ leads to $\mathrm{x} = \boldsymbol{\mu}$. Therefore, the mode of the multivariate Gaussian distribution is equal to $\boldsymbol{\mu}$.

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

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

$$var[x+z] = var[x] + var[z]$$

Suppose that x and z are discrete random variable,

$$\mathrm{E}[x+z] = \sum_i \sum_j (x_i + z_j) f(x_i, z_j) = \sum_i \sum_j x_i f(x_i, z_j) + \sum_i \sum_j z_j f(x_i, z_j)$$

$$= \sum_i x_i f(x_i) + \sum_j z_j f(z_j) = \mathrm{E}[x] + \mathrm{E}[z]$$

$$var[x+z] = \mathrm{E}\big[(x+z) - \mathrm{E}[x+z]\big]^2 = \mathrm{E}[(x+z) - (\mathrm{E}[x] + \mathrm{E}[z])]^2$$

$$= \mathrm{E}[(x - \mathrm{E}[x]) + (z - \mathrm{E}[z])]^2$$

$$= \mathrm{E}[(x - \mathrm{E}[x])^2 + (z - \mathrm{E}[z])^2 + 2(x - \mathrm{E}[x])(z - \mathrm{E}[z])]$$

$$= \mathrm{E}[(x - \mathrm{E}[x])^2] + \mathrm{E}[(z - \mathrm{E}[z])^2] + 2Cov(x, y)$$

Since x and z are independent, $Cov(x, y) = 0$ by problem 1.6. Therefore,

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

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).

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

When taking derivative of this log function with respect to μ,

$$\frac{\partial}{\partial\mu} \ln p(\mathrm{x}|\mu, \sigma^2) = -\frac{1}{\sigma^2} \sum_{n=1}^{N} (x_n - \mu_{ML}) = \sum_{n=1}^{N} x_n - N\mu_{ML} = 0$$

$$\frac{\partial}{\partial \sigma} \ln p(\mathrm{x}|\mu, \sigma^2) = \frac{1}{\sigma^3} \sum_{n=1}^{N} (x_n - \mu_{ML})^2 - \frac{N}{\sigma} = 0$$

To satisfy the equation,

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

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

1.13 Suppose that the variance of a Gaussian is estimated using the result (1.56) but with the maximum likelihood estimates μ_{ML} replaced with the true value μ of the mean. Show that this estimator has the property that its expectation is given by the true variance σ^2.

Expectation of σ_{ML}^2 after replacing μ_{ML} with μ is

$$\mathrm{E}[\sigma_{ML}^2] = \mathrm{E}\left[\frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2\right] = \frac{1}{N} \mathrm{E}\left[\sum_{n=1}^{N} x_n^2 - 2x_n\mu + \mu^2\right] = \frac{1}{N}\left\{\sum_{n=1}^{N} \mathrm{E}[x_n^2] - 2N\mu^2 + N\mu^2\right\}$$

$$= \frac{1}{N}\{N\mathrm{E}[x_n^2] - N\mu^2\} = \mathrm{E}[x_n^2] - \mu^2 = Var[x] = \sigma^2$$