INFR10086 Machine Learning (MLG) Semester 2, 2025/6

Exercises 1

Lecturer: Hao Tang

Exercise 1. Check that the mean of a 1D Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \tag{1}$$

is μ by showing that $\mathbb{E}[x] = \mu$.

One solution is to directly evaluate

$$\mathbb{E}[x] = \int x p(x) dx = \int x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx, \tag{2}$$

perhaps using integration by part. This approach works but is also quite hairy.

One observation is that we only need an additional x in front of the Gaussian density. An ingenious solution is to start with the integral of a Gaussian

$$\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = 1, \tag{3}$$

and take the derivative of μ on both sides. We end up with

$$\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \frac{1}{\sigma^2}(x-\mu) dx = 0. \tag{4}$$

Rearranging the terms, we have

$$\int x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = \int \mu \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx. \tag{5}$$

The left hand side becomes

$$\mathbb{E}[x] = \int x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx, \tag{6}$$

while the right hand side is

$$\int \mu \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = \mu \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = \mu. \tag{7}$$

Exercise 2. Show that the log of a multivariate Gaussian distribution

$$\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{8}$$

is given as

$$-\frac{1}{2}\mathbf{x}^\top \Sigma^{-1}\mathbf{x} + \boldsymbol{\mu}^\top \Sigma^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu} - \frac{1}{2}\log|\Sigma| - \frac{d}{2}\log 2\pi. \tag{9}$$

The key step is expanding the quadratic form

$$\begin{aligned}
(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= (\mathbf{x}-\boldsymbol{\mu})^\top (\Sigma^{-1}\mathbf{x} - \Sigma^{-1}\boldsymbol{\mu}) && (10)\\
&= (\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}\mathbf{x} - (\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}\boldsymbol{\mu} && (11)\\
&= \mathbf{x}^\top \Sigma^{-1}\mathbf{x} - \boldsymbol{\mu}^\top \Sigma^{-1}\mathbf{x} - \mathbf{x}^\top \Sigma^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu} && (12)\\
&= \mathbf{x}^\top \Sigma^{-1}\mathbf{x} - 2\boldsymbol{\mu}^\top \Sigma^{-1}\mathbf{x} + \boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu}. && (13)
\end{aligned}$$

Note that these are all matrix-vector mulplication, and the transpose signs are necessary in the derivation. The rest simply follows from taking the log.

Exercise 3. Show that a covariance matrix $\Sigma = \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top]$ is symmetric and positive semidefinite. A matrix A is positive semidefinite if $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x}$.

To check for symmetry,

$$\begin{aligned}
\Sigma^\top &= \left(\mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top]\right)^\top = \mathbb{E}[((\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top)^\top] = \mathbb{E}[((\mathbf{x}-\boldsymbol{\mu})^\top)^\top(\mathbf{x}-\boldsymbol{\mu})^\top] && (14)\\
&= \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top] = \Sigma. && (15)
\end{aligned}$$

To check for positive semidefiniteness, for any $\mathbf{v}$,

$$\mathbf{v}^\top \Sigma \mathbf{v} = \mathbf{v}^\top \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top]\mathbf{v} = \mathbb{E}[\mathbf{v}^\top(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{v}] = \mathbb{E}\left[\left(\mathbf{v}^\top(\mathbf{x}-\boldsymbol{\mu})\right)^2\right] \geq 0. \tag{16}$$

Exercise 4. In this question, we will look at why the contour of Gaussian distributions consists of ellipses. A general definition of an ellipse (or an ellipsoid in high dimensions) can be written as

$$\left\{\mathbf{x} \in \mathbb{R}^d \,\middle|\, (\mathbf{x}-\mathbf{v})^\top A(\mathbf{x}-\mathbf{v}) = 1\right\}, \tag{17}$$

where $\mathbf{v}$ is where the ellipse is centered and A is a symmetric and positive definite matrix. A matrix A is positive definite if $\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

- An axis-aligned ellipse in 2D can be written as

$$\left\{(x, y) \in \mathbb{R}^2 \,\middle|\, \frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1\right\}, \tag{18}$$

where a and b are the lengths of the two axes and (x_0, y_0) is where the ellipse is centered. Show that equation (18) can be written as equation (17).

- A contour of Gaussian distribution consists of lines where the distribution has the same value. Show that

$$\left\{\mathbf{x} \in \mathbb{R}^d \,\middle|\, \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) = c\right\} \tag{19}$$

is an ellipse for some constant $c > 0$ by rewriting it as equation (17), assuming that the covariance matrix is positive definite.

- We take $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^\top$, $\mathbf{v} = \begin{bmatrix} x_0 & y_0 \end{bmatrix}^\top$ and $A = \begin{bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{bmatrix}$. It's easy to see that A is positive definite. We can then write equation (18) in (17).
- We can take the log on both sides and get

$$(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = (-2)\left(\log c + \frac{d}{2}\log 2\pi + \frac{1}{2}\log|\Sigma|\right). \tag{20}$$

When the covariance matrix is positive definite, its inverse is also positive definite.[1] The right hand side is thus positive, and we can divide it to the left and complete our ellipsoid.

Exercise 5. In a classification setting, where $x \in \mathbb{R}^d$, $y \in \{1, \ldots, K\}$, and K is the number of classes, show that

$$p(y|x) = \frac{p(x|y)p(y)}{\sum_{y'=1}^K p(x|y')p(y')}. \tag{21}$$

This is a straight application of the Bayes rule, where

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_{y'=1}^K p(x|y')p(y')}. \tag{22}$$

Exercise 6. Consider binary classification with the linear classifier

$$y(\mathbf{x}) = \begin{cases} +1 & \text{if } \mathbf{w}^\top\mathbf{x} + b \geq 0 \\ -1 & \text{otherwise} \end{cases} \tag{23}$$

where $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{w}$ is the weight vector, and b is the bias.

- Show that the decision boundary is a straight line when $d = 2$. A line in 2D can be expressed as $y = ax + b$ for some constant $a \in \mathbb{R}$ and $b \in \mathbb{R}$.

[1] When a matrix is positive definite, its eigenvalues are positive. The eigenvalues of A^{-1} are the reciprocal of the eigenvalues of A, so A^{-1} is also positive definite.

- Show that the weight vector $\mathbf{w}$ is a normal vector of the decision boundary.

- The decision boundary can be written as

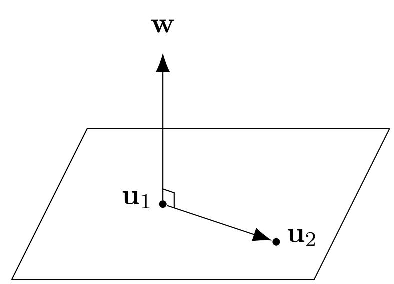
$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}^\top \begin{bmatrix} x \\ y \end{bmatrix} + b = w_1 x + w_2 y + b = 0. \tag{24}$$

  It is now easy to see that

$$y = -\frac{w_1}{w_2} x - \frac{b}{w_2}, \tag{25}$$

  which is the line we are looking for.

- To show that $\mathbf{w}$ is a normal vector of the decision boundary (a plane), the plan is to take any vector on the plane and show that the dot product with $\mathbf{w}$ is 0.

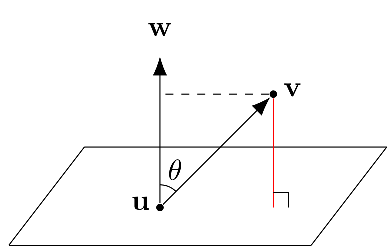


  To get a vector on the plane, we subtract two points $\mathbf{u}_1$ and $\mathbf{u}_2$ on the plane. Next, we compute the dot product

$$\mathbf{w}^\top(\mathbf{u}_2 - \mathbf{u}_1) = \mathbf{w}^\top \mathbf{u}_2 - \mathbf{w}^\top \mathbf{u}_1 = \mathbf{w}^\top \mathbf{u}_2 + b - (\mathbf{w}^\top \mathbf{u}_1 + b) = 0, \tag{26}$$

  where the last equality uses the fact that any point $\mathbf{u}$ on the decision boundary satisfies $\mathbf{w}^\top \mathbf{u} + b = 0$.

Exercise 7. Derive a formula for the Euclidean distance between the origin $(0, 0)$ and a line $y = ax + b$, where a and b are arbitrary constants.



Based on the figure, the length of the red line is the distance between a plane and the point $\mathbf{v}$, where $\mathbf{u}$ is some other point on the plane and $\mathbf{w}$ is the normal vector of the plane. The red line can be described as $\|\mathbf{v}-\mathbf{u}\||\cos\theta|$ where θ is the angle between $\mathbf{v}-\mathbf{u}$ and $\mathbf{w}$. For this question, $\mathbf{v}$ is the origin $(0,0)$; $\mathbf{u}$ is some point on the plane and we can simply take $(0,b)$; $\mathbf{w}$ is the normal vector $(a,-1)$. The distance is then

$$\|\mathbf{v}-\mathbf{u}\||\cos\theta| = \|\mathbf{v}-\mathbf{u}\| \left| \frac{(\mathbf{v}-\mathbf{u})^\top \mathbf{w}}{\|\mathbf{v}-\mathbf{u}\|\|\mathbf{w}\|} \right| = \frac{|(\mathbf{v}-\mathbf{u})^\top \mathbf{w}|}{\|\mathbf{w}\|} \tag{27}$$

$$= \frac{\left| \left[\begin{bmatrix} 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & b \end{bmatrix} \right]^\top \begin{bmatrix} a & -1 \end{bmatrix} \right|}{\sqrt{1+a^2}} = \frac{b}{\sqrt{1+a^2}}. \tag{28}$$

Exercise 8. Consider the 2D case for the linear classifier in equation (23). Suppose that the points $(-2,-3)$ and $(4,1)$ are on the decision bounary and that the point $(2,-3)$ lies in the -1 class region. Find the parameters $(\mathbf{w}, b)$ of the classifier.

From equation (24), we see that (a,b) is the normal vector if we write the line as $ax+by+c=0$. If the line passes through $(-2,-3)$ and $(4,1)$, then the normal vector should be perpendicular to $(4,1)-(-2,-3)=(6,4)$. We can simply choose $(2,-3)$ to be our normal vector, and the rest is to figure out what c is in $2x-3y+c=0$. The line passes through $(4,1)$, or $2\cdot 4-3\cdot 1+c=0$, so c is -5 and the line is $2x-3y-5=0$. We also have the choice of writing the line as $-2x+3y+5=0$. To decide whether we want to negate the left hand side or not, we can choose the version such that $(2,-3)$ is on the negative side. Because $-2\cdot 2+3\cdot(-3)+5\leq 0$, we choose $-2x+3y+5=0$ as our decision boundary. In other words, $\mathbf{w}=(-2,3)$ and $b=5$.