# Maximum likelihood estimates

i) Derivative with respect to $\mu$:

$$\frac{\partial}{\partial \mu} \ln p(x|\mu, \sigma^2) = \frac{1}{\sigma^2} \sum (x_n - \mu)$$

Setting it equal to 0, we get:

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

ii) Derivative with respect to $\sigma^2$:

$$\frac{\partial}{\partial \sigma^2} \ln p(x|\mu_{ML}, \sigma^2) = \frac{1}{2\sigma^4} \sum (x_n - \mu_{ML})^2 - \frac{N}{2\sigma^2}$$

Setting it equal to 0, we get:

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

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

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

# True variance

Assume that $\sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2$.

Then, we have:

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

$$= \frac{1}{N}\sum_{n=1}^{N} E[x_n^2 - 2x_n\mu + \mu^2]$$

$$= \frac{1}{N}\sum_{n=1}^{N} E[x_n^2] - \frac{2\mu}{N}\sum_{n=1}^{N} E[x_n] + \mu^2$$

Remember that $Var(x) = E[x^2] - (E[x])^2$

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

$$= \sigma^2 + \mu^2 - 2\mu^2 + \mu^2$$

$$= \sigma^2$$

# Symmetry

First of all, we can write:

$$w_{ij} = w_{ij} + \frac{w_{ji}}{2} - \frac{w_{ji}}{2} = \frac{w_{ij} + w_{ji}}{2} + \frac{w_{ij} - w_{ji}}{2}$$

We set $w_{ij}^S = \frac{w_{ij} + w_{ji}}{2} = \frac{w_{ji} + w_{ij}}{2} = w_{ji}^S$

and $w_{ij}^A = \frac{w_{ij} - w_{ji}}{2} = -\frac{w_{ji} - w_{ij}}{2} = -w_{ji}^A$

i) We have:

$$\begin{aligned}
\sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij} x_i x_j &= x^T W x \\
&= x^T (W^S + W^A) x \\
&= x^T W^S x + x^T W^A x \\
&= x^T W^S x + \frac{1}{2} x^T W^A x + \frac{1}{2} x^T (W^A x) \\
&= x^T W^S x + \frac{1}{2} x^T W^A x + \frac{1}{2} (W^A x)^T (x^T)^T \quad \text{Symmetry of the dot product} \\
&= x^T W^S x + \frac{1}{2} x^T W^A x + \frac{1}{2} x^T (W^A)^T x \\
&= x^T W^S x + \frac{1}{2} x^T W^A x - \frac{1}{2} x^T W^A x \quad W^A \text{ is anti-symmetric} \\
&= x^T W^S x \\
&= \sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij}^S x_i x_j
\end{aligned}$$

ii) In a symmetric $D \times D$ matrix, setting the value of $w_{ij}$ gives us $w_{ji} = +w_{ij}$. Therefore, we can set independently the values above (or below) the diagonal and the values of the diagonal itself.

The diagonal has $D$ elements and, from the remaining $D^2$ elements, we can set half of the values independently. Therefore, the number of independent parameters is given by:

$$D + \frac{D^2 - D}{2} = \frac{2D + D^2 - D}{2} = \frac{D^2 + D}{2} = \frac{D(D+1)}{2}$$

# Misclassification bound

Since $a \geq 0$, we have $a \leq b \Rightarrow a^2 \leq ab \Rightarrow a \leq \sqrt{ab}$

Let $R_1, R_2$ be the two decision regions.

Then, $p(\text{mistake}) = p(x \in R_1, C_2) + p(x \in R_2, C_1)$

$$= \int_{R_1} p(x, C_2)\,dx + \int_{R_2} p(x, C_1)\,dx$$

$$= \int_{R_1} p(C_2|x)\,p(x)\,dx + \int_{R_2} p(C_1|x)\,p(x)\,dx$$

The regions that minimize the probability of misclassification are such that $p(C_1|x) \geq p(C_2|x)$ for all $x \in R_1$ and $p(C_2|x) \geq p(C_1|x)$ for all $x \in R_2$. Using the first result, we get:

$$p(\text{mistake}) \leq \int_{R_1} \sqrt{p(C_2|x)\,p(C_1|x)}\,p(x)\,dx + \int_{R_2} \sqrt{p(C_1|x)\,p(C_2|x)}\,p(x)\,dx$$

$$= \int_{R_1} \sqrt{p(x, C_1)\,p(x, C_2)}\,dx + \int_{R_2} \sqrt{p(x, C_1)\,p(x, C_2)}\,dx$$

$$= \int \sqrt{p(x, C_1)\,p(x, C_2)}\,dx$$

# Minimal loss (i)

i) We have that:

$$\sum_k L_{kj} p(C_k|x) = \sum_k (1 - I_{kj}) p(C_k|x)$$
$$= \sum_k p(C_k|x) - \sum_k I_{kj} p(C_k|x)$$
$$= 1 - \sum_k I_{kj} p(C_k|x)$$
$$= 1 - p(C_j|x)$$

Therefore, we get:

$$\underset{j}{\arg\min} \sum_k L_{kj} p(C_k|x) = \underset{j}{\arg\min} \left[1 - p(C_j|x)\right]$$
$$= \underset{j}{\arg\max}\ p(C_j|x)$$

ii) The loss matrix $L$ has the form:

$$L = \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 0 \end{bmatrix}$$

Let $k$ be the real class. If we predict a class $j=k$, the loss is equal to 0. Otherwise, for any predicted class $j \neq k$, we get loss equal to 1.

# Minimal loss (ii)

The expected loss is:

$$E[L] = \sum_k \sum_j \int_{R_j} L_{kj}\, p(x, C_k)\, dx$$

$$= \sum_k \sum_j \int_{R_j} L_{kj}\, p(x|C_k) \cdot p(C_k)\, dx$$

we set $L'_{kj} = L_{kj} \cdot p(C_k)$

$$= \sum_k \sum_j \int_{R_j} L'_{kj} \cdot p(x|C_k)\, dx$$

$$= \sum_j \int_{R_j} \sum_k L'_{kj} \cdot p(x|C_k)\, dx$$

Therefore, to minimize it, we have to design the regions $R_j$ such that every $x$ is assigned to the class $C_j$ that minimizes the quantity $\sum_k L'_{kj} \cdot p(x|C_k)$. To see this, let $R_1^*, R_2^*, \dots$ be the decision regions giving the minimum expected loss $L^*$. Now, assume there is a subset $R_e^* \subseteq R_\ell^*$ where $R_\ell^*$ is the decision region of class $C_\ell$ and, for all $x \in R_e^*$, it holds: $\ell \neq i = \arg\min_j \sum_k L'_{kj} \cdot p(x|C_k)$.

Then: $L^* = \underbrace{\sum_{j \neq l, i} \int_{R_j^*} \sum_k L'_{kj} \cdot p(x|C_k)\,dx}_{\text{Let's call that } L_0 \text{ for brevity.}} + \sum_{j \in \{l, i\}} \int_{R_j^*} \sum_k L'_{kj} \cdot p(x|C_k)\,dx$

$$= L_0 + \int_{R_e^*} \sum_k L'_{kl} \cdot p(x|C_k)\,dx + \int_{R_l^* \setminus R_e^*} \sum_k L'_{kl} \cdot p(x|C_k)\,dx + \int_{R_i^*} \sum_k L'_{ki} \cdot p(x|C_k)\,dx$$

Let's consider different decision regions $\tilde{R}_1, \tilde{R}_2, \ldots$ such that $\tilde{R}_i = R_i^* \cup R_e^*$, $\tilde{R}_l = R_l^* \setminus R_e^*$ and $\tilde{R}_j = R_j^*$ for all $j \neq i, l$.

Then, it is easy to see that the expected loss is:

$$\tilde{L} = L_0 + \sum_{j \in \{l, i\}} \int_{\tilde{R}_j} \sum_k L'_{kj} \cdot p(x|C_k)\,dx$$

$$= L_0 + \int_{R_e^*} \sum_k L'_{ki} \cdot p(x|C_k)\,dx + \int_{R_i^*} \sum_k L'_{ki} \cdot p(x|C_k)\,dx + \int_{R_l^* \setminus R_e^*} \sum_k L'_{kl} \cdot p(x|C_k)\,dx$$

Therefore, we have

$$\tilde{L} - L^* = \int_{R_e^*} \left[ \sum_k L'_{ki} \cdot p(x|C_k) - \sum_k L'_{kl} \cdot p(x|C_k) \right] dx < 0,$$

which is a contradiction.

Recall that $l \neq i = \arg\min_j \sum_k L'_{kj} \cdot p(x|C_k)$

# Theory (Calculus of variations)

Assume we have a quantity $Q(x, y(x)) = \int G(x, y(x), \dot{y}(x))dx$ and we want to find the function $y(x)$ that maximizes the given quantity. To do that, we define a functional $J[y] = \int G(x, y(x), \dot{y}(x))dx$ and we set its functional derivative $\frac{\delta J}{\delta y}$ to zero.

- **x, y real:** $\frac{\delta J}{\delta y} = 0 \Leftrightarrow \frac{\partial G}{\partial y} - \frac{d}{dx}\frac{\partial G}{\partial \dot{y}} = 0$ where $\frac{\partial G}{\partial y}, \frac{\partial G}{\partial \dot{y}}$ denote the derivatives of $G$ with respect to $y$ and $\dot{y}$ and we can think of $y, \dot{y}$ as independent variables. For example, $G(x, y(x), \dot{y}(x)) = y^2(x) + \dot{y}^3(x)$. Then, $\frac{\partial G}{\partial y} = 2y(x)$ and $\frac{\partial G}{\partial \dot{y}} = 3\dot{y}^2(x)$.

- **x vector, y real:** $Q(x, y(x)) = \int G(x, y(x), \nabla y(x))dx$, $x$ 3-dimensional

$$\frac{\delta J}{\delta y} = 0 \Leftrightarrow \frac{\partial G}{\partial y} - \nabla \cdot \frac{\partial G}{\partial \nabla y} = 0$$

Example: $G = y^2(x) + \nabla y(x) \cdot \nabla y(x) = y^2(x) + \left(\frac{\partial y}{\partial x_1}\right)^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 + \left(\frac{\partial y}{\partial x_3}\right)^2$

$$\frac{\partial G}{\partial y} = 2y(x), \quad \frac{\partial G}{\partial \nabla y} = \begin{bmatrix} 2\frac{\partial y}{\partial x_1} \\ 2\frac{\partial y}{\partial x_2} \\ 2\frac{\partial y}{\partial x_3} \end{bmatrix} = 2\nabla y(x), \quad \nabla \cdot \frac{\partial G}{\partial \nabla y} = 2\nabla^2 y(x) = 2\sum_{i=1}^{3} \frac{\partial^2 y}{\partial x_i^2}$$

- x vector
  y vector: $J[y] = J[y_1, y_2, \dots, y_m]$

$$= Q(x, y_1(x), y_2(x), \dots, y_m(x))$$

$$= \int G(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m, y_{1,1}, \dots, y_{1,n}, \dots$$
$$\dots, y_{m,1}, \dots, y_{m,n})\,dx$$

where $y_{i,j} = \frac{\partial y_i}{\partial x_j}$

$$\frac{\delta J}{\delta y} = 0 \iff$$

$$\frac{\partial G}{\partial y_i} - \sum_{j=1}^{n} \frac{\partial}{\partial x_j}\left(\frac{\partial G}{\partial y_{i,j}}\right) = 0 \quad \text{for } i = 1, 2, \dots, m$$

Example in the exercise (Targets).

# Targets

Consider the functional

$$J[y] = \int\int \|y(x) - t\|^2 p(x,t)\, dt\, dx = E\left[L(t, y(x))\right]$$

and let $G(x, y(x)) = \int \|y(x) - t\|^2 p(x,t)\, dt.$

To find the function $y(x) = [y_1(x), \dots, y_m(x)]$ we use:

$$\frac{\partial G}{\partial y_i} - \sum_{j=1}^{n} \frac{\partial}{\partial x_j}\left(\frac{\partial G}{\partial y_{i,j}}\right) = 0 \text{ for } i = 1, \dots, m \iff$$

$$\frac{\partial G}{\partial y_i} = 0 \quad \text{for } i = 1, \dots, m \iff$$

$$\int \frac{\partial}{\partial y_i}\left[\|y(x) - t\|^2\right] p(x,t)\, dt = 0 \quad \text{for } i = 1, \dots, m \iff$$

$$\int 2(y_i(x) - t_i)\, p(x,t)\, dt = 0 \quad \text{for } i = 1, \dots, m \iff$$

$$y_i(x) \cdot \int p(x,t)\, dt = \int t_i \cdot p(x,t)\, dt \quad \text{for } i = 1, \dots, m \iff$$

$$y(x) = \frac{\int t \cdot p(x,t)\, dt}{p(x)} = \int t \cdot p(t|x)\, dt = E_t[t|x]$$

# Regression

i) Consider the functional

$$J[y] = \iint |y(x) - t|^q p(x,t)\, dt\, dx = E\left[L(t, y(x))\right]$$

and let $G(x, y(x)) = \int |y(x) - t|^q p(x,t)\, dt$.

To find the function $y(x)$ that minimizes the loss, we use:

$$\frac{\partial G}{\partial y} - \frac{d}{dx}\frac{\partial G}{\partial \dot{y}} = 0 \Leftrightarrow$$

$$\frac{\partial G}{\partial y} = 0 \Leftrightarrow$$

$$q \int [y(x) - t]^{q-1} \cdot \operatorname{sign}(y(x) - t) \cdot p(x,t)\, dt = 0 \Leftrightarrow$$

$$\int_{-\infty}^{y(x)} [y(x) - t]^{q-1} \cdot p(t|x)\, dt = \int_{y(x)}^{\infty} [y(x) - t]^{q-1} \cdot p(t|x)\, dt$$

ii) When $q = 1$, we get:

$$\int_{-\infty}^{y(x)} p(t|x)\, dt = \int_{y(x)}^{\infty} p(t|x)\, dt$$ and $y(x)$ is the conditional median.

iii) We have $E[L_q] = \iint |y(x)-t|^q p(x,t)\,dx\,dt$

$$= \int \left[ \int |y(x)-t|^q p(t|x)\,dt \right] p(x)\,dx.$$

Therefore, the expected loss is minimized if, for each $x$ independently, we choose the value $y(x)$ that minimizes $\int |y(x)-t|^q p(t|x)\,dt$.

We can see that, when $q \to 0$, the quantity $|y(x)-t|^q$ equals $1$ when $t \neq y(x)$ and gets value $0$ in an infinitesimally small region around $t = y(x)$. Therefore, the previous integral is minimized when we set $y(x) = t$ at the point $t$ that maximizes the quantity $p(t|x)$, i.e., the conditional mode.

# Decision boundary

i) It is $P(\text{error}) = P(\text{error}|C_1)\cdot P(C_1) + P(\text{error}|C_2)\cdot P(C_2)$

$$= P(x \leq \theta | C_1)\cdot P(C_1) + P(x > \theta | C_2)\cdot P(C_2)$$

$$= P(C_1)\int_{-\infty}^{\theta} p(x|C_1)dx + P(C_2)\int_{\theta}^{\infty} p(x|C_2)\,dx$$

ii) The second term can be rewritten as:

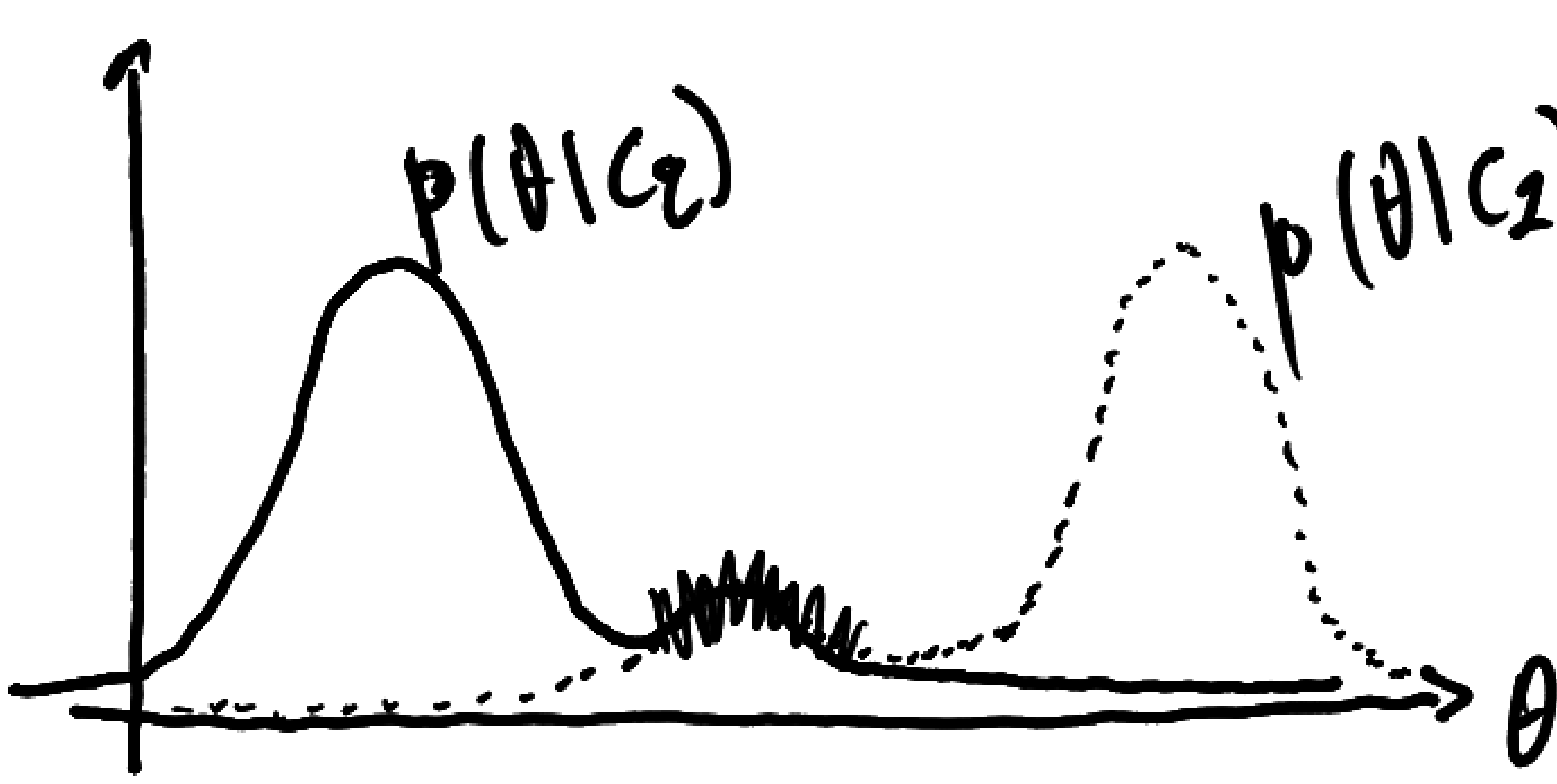
$$P(C_2)\cdot\left[\int_{-\infty}^{\infty} p(x|C_2)dx - \int_{-\infty}^{\theta} p(x|C_2)dx\right] = P(C_2) - P(C_2)\cdot\int_{-\infty}^{\theta} p(x|C_2)dx$$

Therefore, $\dfrac{dP(\text{error})}{d\theta} = 0 \Leftrightarrow$

$$P(C_1)\cdot p(\theta|C_1) - P(C_2)\cdot p(\theta|C_2) = 0 \Leftrightarrow$$

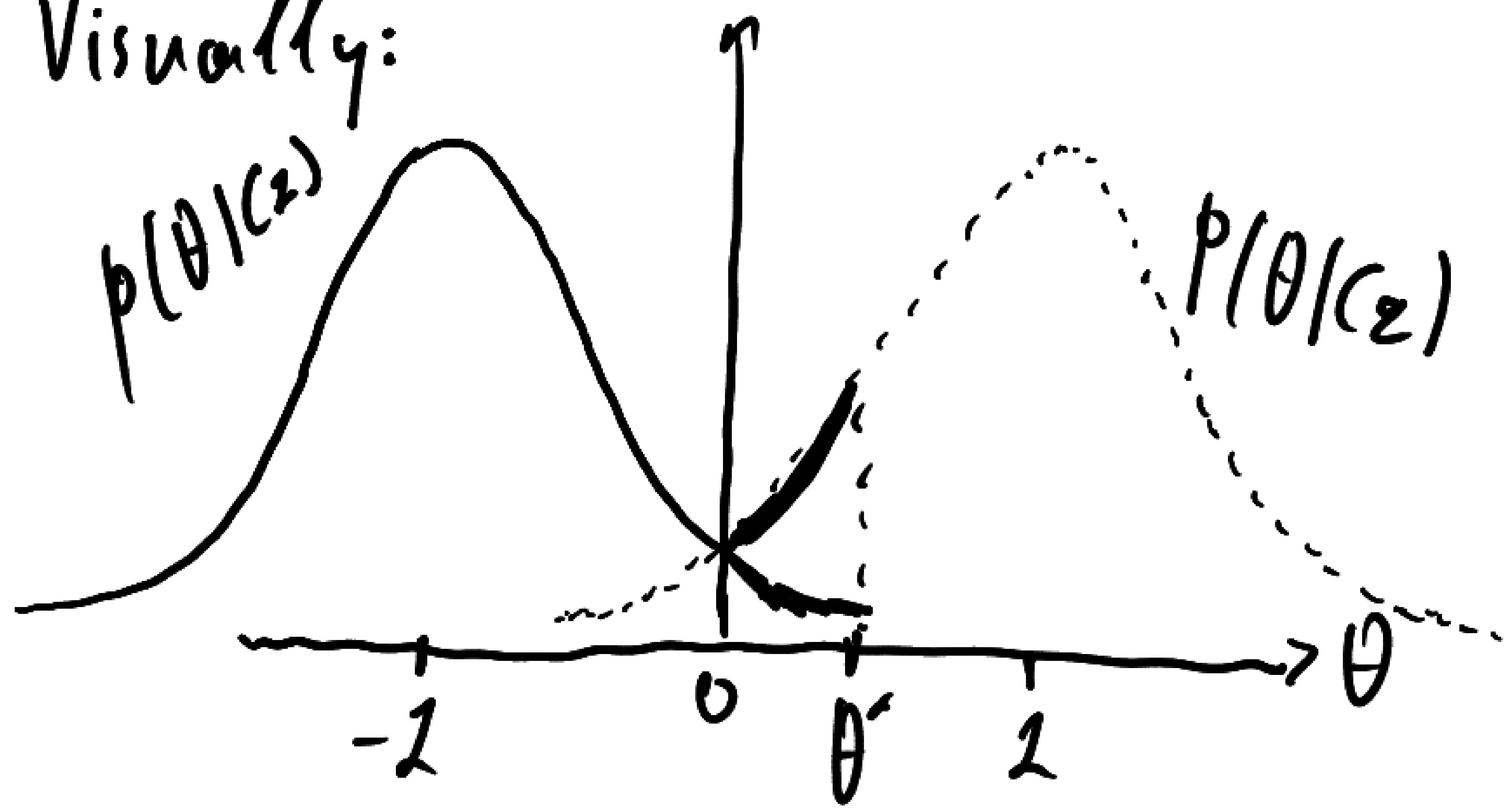
$$P(C_1)\cdot p(\theta|C_1) = P(C_2)\cdot p(\theta|C_2)$$

iii) It does not uniquely identify $\theta$. It could be $P(C_1) = P(C_2)$. Then, it suffices to find a range of $\theta$ where the conditionals match.



iv) Assume $P(C_1) = P(C_2) = \frac{1}{2}$ and $\theta | C_1 \sim \mathcal{N}(-1,1)$, $\theta | C_2 \sim \mathcal{N}(1,1)$.

Visually:




The condition gives $\theta = 0$. Then,:

$$e^* = P(\text{error}) = \frac{1}{2} \int_{-\infty}^{0} p(x|C_1)\,dx + \frac{1}{2} \int_{0}^{\infty} p(x|C_2)\,dx$$

For every $\theta' > 0$, we have:

$$e = P(\text{error}) = \frac{1}{2} \int_{-\infty}^{0} p(x|C_1)\,dx + \frac{1}{2} \int_{0}^{\theta'} p(x|C_1)\,dx + \frac{1}{2} \int_{0}^{\infty} p(x|C_2)\,dx - \frac{1}{2} \int_{0}^{\theta'} p(x|C_2)\,dx$$

$$= e^* + \frac{1}{2} \int_{0}^{\theta'} \left[ p(x|C_1) - p(x|C_2) \right] dx$$

$$< e^*$$

Similarly, for $\theta' < 0$:

$$e = e^* + \frac{1}{2} \int_{\theta'}^{0} \left[ p(x|C_2) - p(x|C_1) \right] dx < e^*.$$

Optional: What is the optimal $\theta$ in the previous example?

We have that:

$$P(\text{error}) = P(X \leq \theta \mid C_1) \cdot \frac{1}{2} + P(X > \theta \mid C_2) \cdot \frac{1}{2}$$

$$= P\left(Z \leq \frac{\theta - (-1)}{1}\right) \cdot \frac{1}{2} + P\left(Z > \frac{\theta - 1}{1}\right) \cdot \frac{1}{2}$$ where $Z \sim \mathcal{N}(0,1)$

$$= P(Z \leq \theta + 1) \cdot \frac{1}{2} + \left[1 - P(Z \leq \theta - 1)\right] \cdot \frac{1}{2}$$

$$= \frac{1}{2} + \frac{1}{2}\left[P(Z \leq \theta + 1) - P(Z \leq \theta - 1)\right]$$

$$= \frac{1}{2} + \frac{1}{2} P(\theta - 1 \leq Z \leq \theta + 1)$$

For every $\theta \in \mathbb{R}$, it is easy to see that $P(\text{error}) > \frac{1}{2}$.
Instead, we get $\lim_{\theta \to -\infty} P(\text{error}) = \lim_{\theta \to \infty} P(\text{error}) = \frac{1}{2}$.
In the limit, we either classify all samples as $C_1$ (and mis-classify all samples of $C_2$) or the opposite.
Because $P(C_1) = P(C_2) = \frac{1}{2}$, we get $P(\text{error}) = \frac{1}{2}$.

# At the limit

i) The minimum-error-rate decision rule returns $C_1$ if $P(C_1|x) > P(C_2|x)$ and $C_2$ otherwise.

We have:

$$P(C_1|x) > P(C_2|x) \iff$$

$$\frac{P(x|C_1)\cdot P(C_1)}{P(x)} > \frac{P(x|C_2)\cdot P(C_2)}{P(x)} \iff \quad P(C_1) = P(C_2) = \frac{1}{2}$$

$$P(x|C_1) > P(x|C_2) \iff$$

$$\prod_{i=1}^{d} P(x_i|C_1) > \prod_{i=1}^{d} P(x_i|C_2) \iff$$

$$\sum_{i=1}^{d} \log P(x_i|C_1) > \sum_{i=1}^{d} \log P(x_i|C_2) \iff$$

$$\sum_{i=1}^{d} \log\left[p^{x_i}(1-p)^{(1-x_i)}\right] > \sum_{i=1}^{d} \log\left[(1-p)^{x_i} p^{(1-x_i)}\right] \iff$$

$$\sum_{i=1}^{d} x_i \log p + \sum_{i=1}^{d}(1-x_i)\log(1-p) > \sum_{i=1}^{d} x_i \log(1-p) + \sum_{i=1}^{d}(1-x_i)\log p \iff$$

$$\log p \cdot \sum_{i=1}^{d} x_i + d\cdot\log(1-p) - \log(1-p)\cdot\sum_{i=1}^{d} x_i > \log(1-p)\cdot\sum_{i=1}^{d} x_i + d\log p - \log p \cdot \sum_{i=1}^{d} x_i \iff$$

$$\sum_{i=1}^{d} x_i \cdot\left[2\log p - 2\log(1-p)\right] > d\left[\log p - \log(1-p)\right] \iff$$

$$\sum_{i=1}^{d} x_i > \frac{d}{2}.$$

ii) The probability of error is:

$$P_e(d,p) = P(\text{error} \mid C_1)\cdot P(C_1) + P(\text{error} \mid C_2)\cdot P(C_2)$$

$$= P\left(\sum_{i=1}^{d} x_i \le \frac{d}{2} \mid C_1\right)\cdot \frac{1}{2} + P\left(\sum_{i=1}^{d} x_i > \frac{d}{2} \mid C_2\right)\cdot \frac{1}{2}$$

$$= P\left(\sum_{i=1}^{d} x_i \le \frac{d-1}{2} \mid C_1\right)\cdot \frac{1}{2} + P\left(\sum_{i=1}^{d} x_i \ge \frac{d+1}{2} \mid C_2\right)\cdot \frac{1}{2}$$

Recall that $d$ is odd

$$= \frac{1}{2}\sum_{k=0}^{(d-1)/2} \binom{d}{k} p^k (1-p)^{(d-k)} + \frac{1}{2}\sum_{k=(d+1)/2}^{d} \binom{d}{k} (1-p)^k \cdot p^{(d-k)}$$

In the second term, we set $j = d-k$ and we get:

$$P_e(d,p) = \frac{1}{2}\sum_{k=0}^{(d-1)/2} \binom{d}{k} p^k (1-p)^{d-k} + \frac{1}{2}\sum_{j=0}^{(d-1)/2} \binom{d}{d-j} p^j (1-p)^{d-j}$$

Recall that $\binom{d}{d-j} = \binom{d}{j}$

$$= \sum_{k=0}^{(d-1)/2} \binom{d}{k} p^k (1-p)^{d-k}$$

iii)

$$\lim_{p\to 1/2} P_e(d,p) = \sum_{k=0}^{(d-1)/2} \binom{d}{k} \lim_{p\to 1/2} p^k (1-p)^{d-k}$$

$$= \sum_{k=0}^{(d-1)/2} \binom{d}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{d-k}$$

$$= \left(\frac{1}{2}\right)^d \cdot \sum_{k=0}^{(d-1)/2} \binom{d}{k}$$

We have that:

$$\sum_{k=0}^{d}\binom{d}{k} = 2^d \Rightarrow$$

Recall that d is odd

$$\sum_{k=0}^{(d-1)/2}\binom{d}{k} + \sum_{k=(d+1)/2}^{d}\binom{d}{k} = 2^d$$

We set $j = d-k$ and similarly as before, we get:

$$2\cdot\sum_{k=0}^{(d-1)/2}\binom{d}{k} = 2^d \Rightarrow \sum_{k=0}^{(d-1)/2}\binom{d}{k} = 2^{d-1}$$

Therefore, $\lim\limits_{p\to 1/2} P_e(d,p) = \left(\frac{1}{2}\right)^d \cdot 2^{d-1} = \frac{1}{2}$.

iv) We can rewrite $P_e(d,p)$ as $P(0 \leq X \leq (d-1)/2)$ where $X$ is a random variable following $\mathrm{Bin}(d,p)$.
As $d \to \infty$, $X$ converges in distribution to a normal distribution $\mathcal{N}(dp, dp(1-p))$. Therefore:

$$\lim_{d\to\infty} P_e(d,p) = \lim_{d\to\infty} P(0 \leq X \leq (d-1)/2)$$

$$= \lim_{d\to\infty} P\left(\frac{0-dp}{\sqrt{dp(1-p)}} \leq Z \leq \frac{(d-1)/2 - dp}{\sqrt{dp(1-p)}}\right)$$

$$= \lim_{d\to\infty} P\left(-\sqrt{d}\,\frac{\sqrt{p}}{\sqrt{1-p}} \leq Z \leq \frac{\sqrt{d}\,(1/2-p) - \frac{1}{2\sqrt{d}}}{\sqrt{p(1-p)}}\right)$$

where $Z \sim \mathcal{N}(0,1)$.

It is easy to see that the left hand side converges to $-\infty$ and, since $p > \frac{1}{2} \Rightarrow \sqrt{d}\,(1/2 - p) < 0$, the right hand side also converges to $-\infty$.

Therefore $\lim_{d \to \infty} P_e(d, p) = P(-\infty \leq Z \leq -\infty) = 0.$