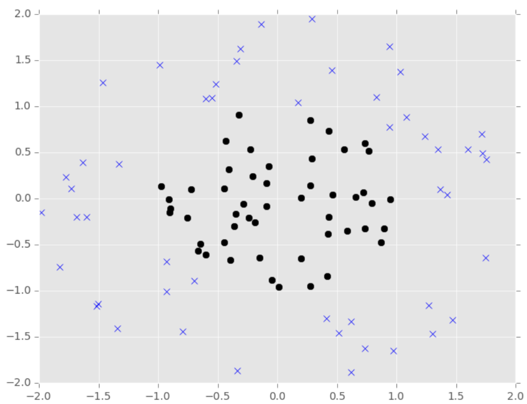
# CS 189: Introduction to Machine Learning, Discussion 6

1. Circular Distributions

Consider the following dataset where each point $x_n = (x_{1,n}, x_{2,n})$ is sampled *iid* and




uniformly at random from two equiprobable (each equally likely) classes, a disk of radius 1 ($y_n = 1$) and a ring from 1 to 2 ($y_n = -1$).

Qualitatively estimate and sketch the following quantities

- The class conditional density $P(X \mid Y)$
- The density of $X$
- The conditional density of $P(Y \mid X)$
- The Bayes Classifier
- The Bayes Risk

**Solution:**

- $$f_{X|Y=1} = \begin{cases} \frac{1}{\pi} & \|X\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$
  $$f_{X|Y=-1} = \begin{cases} \frac{1}{3\pi} & 1 \leq \|X\|_2 \leq 2 \\ 0 & o.w. \end{cases}$$
- $$f_X = \frac{1}{2}\left(f_{X|Y=1} + f_{X|Y=-1}\right)$$

- $$\mathbb{P}[Y=1 \mid X] = \begin{cases} 1 & \|X\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$

  $$\mathbb{P}[Y=-1 \mid X] = \begin{cases} 1 & 1 \leq \|X\|_2 \leq 2 \\ 0 & o.w. \end{cases}$$

- Bayes Classifier: the circle $x_1^2 + x_2^2 = 1$
- The Bayes Risk $= 0$

Now suppose you're given only one of the axes, namely $x_{1,n}$, re-estimate these quantities.

**Solution:**

- $$f_{X|Y=1} = \begin{cases} \frac{2}{\pi}\sqrt{1-x_1^2} & \|x_1\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$

  $$f_{X|Y=-1} = \begin{cases} \frac{2}{3\pi}\sqrt{4-x_1^2} & 1 \leq \|x_1\|_2 \leq 2 \\ \frac{2}{3\pi}\left(\sqrt{4-x_1^2} - \sqrt{1-x_1^2}\right) & \|x_1\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$

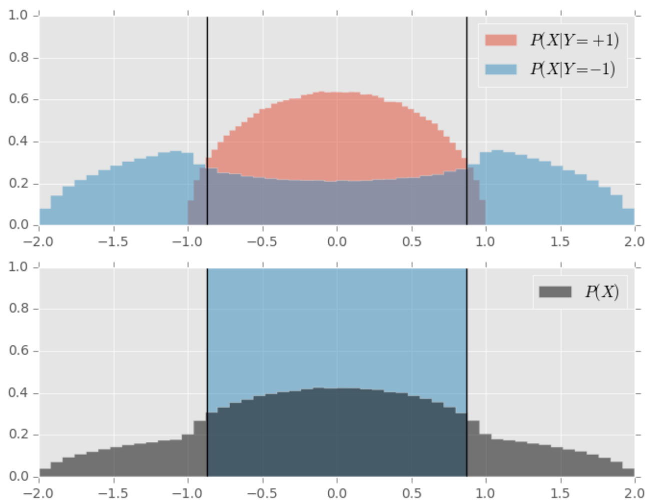
- $$\begin{aligned} f_X &= \frac{1}{2}\left(f_{X|Y=1} + f_{X|Y=-1}\right) \\ &= \begin{cases} \frac{1}{3\pi}\sqrt{4-x_1^2} & 1 \leq \|x_1\|_2 \leq 2 \\ \frac{1}{3\pi}\left(\sqrt{4-x_1^2} + 2\sqrt{1-x_1^2}\right) & \|x_1\|_2 \leq 1 \\ 0 & o.w. \end{cases} \end{aligned}$$

- $$\mathbb{P}[Y=1 \mid X] = \begin{cases} 1 - \frac{\sqrt{4-x_1^2}-\sqrt{1-x_1^2}}{\sqrt{4-x_1^2}+2\sqrt{1-x_1^2}} & \|x_1\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$

  $$\mathbb{P}[Y=-1 \mid X] = \begin{cases} 1 & 1 \leq \|x_1\|_2 \leq 2 \\ \frac{\sqrt{4-x_1^2}-\sqrt{1-x_1^2}}{\sqrt{4-x_1^2}+2\sqrt{1-x_1^2}} & \|x_1\|_2 \leq 1 \\ 0 & o.w. \end{cases}$$

- Bayes Classifier: $\hat{y} = 1$ if $\|x_1\|_2 \leq \sqrt{\frac{4}{5}} = 0.894$
- The Bayes Risk

$$R^* = 2\left(\int_0^{\sqrt{4/5}} \mathbb{P}\left[Y = -1 \mid X = x\right] f_X(x)dx + \int_{\sqrt{4/5}}^1 \mathbb{P}\left[Y = 1 \mid X = x\right] f_X(x)dx\right)$$
$$= 2(0.10 + 0.01) = 0.22$$



Review: Multivariate Gaussian

(a) **True/False** If $X_1$ and $X_2$ are both normally distributed and independent, then $(X_1, X_2)$ must have multivariate normal distribution.

**Solution:** True.
Since $X_1$ and $X_2$ are independent with each other, so we have

$$\begin{aligned} P(X_1, X_2) &= P(X_1)P(X_2) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(X_1-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(X_2-\mu_2)^2}{2\sigma_2^2}} \\ &\sim \mathcal{N}\left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right) \end{aligned}$$

(b) **True/False** If $(X_1, X_2)$ has multivariate normal distribution, then $X_1$ and $X_2$ are independent.

**Solution:** False. If the off diagonal elements of the covariance matrix $\Sigma$ are not zeros, it means $cov(X_1, X_2) \neq 0$. Then they are not independent.

(c) **Affine Transformations** Suppose $\mathbf{X} = [X_1, X_2, \cdots, X_n]^T$ is a n-dimensional random vector which has multivariate Gaussian distribution. Given $\mathbf{X} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ and $\mathbf{y} = \mathbf{c} + \mathbf{B}\mathbf{X}$ is an affine transformation of $\mathbf{X}$, where c is an $M \times 1$ vector of constants and $\mathbf{B}$ is a constant $M \times N$ matrix, what is the expectation and variance of $\mathbf{y}$?

**Solution:**

$$\begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_M \end{pmatrix} + \begin{pmatrix} \mathbf{B_1} \\ \vdots \\ \mathbf{B_M} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_M \end{pmatrix} + \begin{pmatrix} \mathbf{B_1 X} \\ \vdots \\ \mathbf{B_M X} \end{pmatrix} = \begin{pmatrix} c_1 + \mathbf{B_1 X} \\ \vdots \\ c_M + \mathbf{B_M X} \end{pmatrix}$$

So, $E\mathbf{y_i} = E(c_i + \mathbf{B_i X}) = c_i + E(\mathbf{B_i X})$ where $\mathbf{B_i}$ is 1xN vector and $\mathbf{X}$ is Nx1 vector and$\mathbf{B_i X} = \sum_{j=1}^{N} B_{ij} X_j$ for $i = \{1, \cdots, M\}$.
$E(\mathbf{B_i X}) = \sum_{j=1}^{N} B_{ij} E(X_j) = \mathbf{B_i} E(\mathbf{X}) = \mathbf{B_i}\mu$
So, we have $E(y_i) = c_i + \mathbf{B_i}\mu$. Therefore, $E(\mathbf{y}) = \mathbf{c} + \mathbf{B}\mu$

Similarly,

$$\begin{aligned} Var(\mathbf{y}) &= E[(\mathbf{Y}-\mathbf{EY})(\mathbf{Y}-\mathbf{EY})^{\mathbf{T}}] = E[(\mathbf{BX}-\mathbf{B}\mu)(\mathbf{BX}-\mathbf{B}\mu)^{\mathbf{T}}] \\ &= E(\mathbf{B}(\mathbf{X}-\mu)(\mathbf{X}-\mu)^{\mathbf{T}}\mathbf{B}) = \mathbf{B}E[(\mathbf{X}-\mathbf{EX})(\mathbf{X}-\mathbf{EX})^{\mathbf{T}}]\mathbf{B}^{\mathbf{T}} \\ &= \mathbf{B}\ \mathbf{Var}(\mathbf{X})\ \mathbf{B}^{\mathbf{T}} = \mathbf{B}\ \mathbf{\Sigma}\ \mathbf{B}^{\mathbf{T}} \end{aligned}$$

Actually, $\mathbf{y}$ must have a multivariate normal distribution with expected value $\mathbf{c}+\mathbf{B}\mu$ and variance $\mathbf{B\Sigma B^T}$ i.e., $\mathbf{y} \sim \mathcal{N}\left(\mathbf{c}+\mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}^{\mathrm{T}}\right)$. The proof requires some advanced linear algebra and probability theory. If interested, please see the Appendix A.2 of multivariate gaussian worksheet on piazza.

3. Bayesian Decision Theory: Case Study - We're going fishing!

We want to design an automated fishing system that captures fish, classifies them, and sends them off to two different companies, Salmonites, Inc., and Seabass, Inc. For some reason we only ever catch salmon and seabass. Salmonites, Inc. wants salmon, and Seabass, Inc. wants seabass. Given only the weights of the fish we catch, we want to figure out what type of fish it is using machine learning!

Let us assume that the weight of both seabass and salmon are both normally distributed (univariate Gaussian), given by the p.d.f.:

$$P(x|\mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$

We are given this data:

Data for salmon: $\{3, 4, 5, 6, 7\}$

Data for seabass: $\{5, 6, 7, 8, 9, 7+\sqrt{2}, 7-\sqrt{2}\}$

When we classify seabass incorrectly, it gets sent to Salmonites, Inc. who won't pay us for the wrong fish and sells it themselves. When we classify salmon incorrectly, it gets sent to SeaBass, Inc., who is nice and returns our fish. This situation gives rise to this loss matrix:

Predicted:

| Truth: | | **salmon** | **seabass** |
|---|---|---|---|
| | **salmon** | 0 | 1 |
| | **seabass** | 2 | 0 |

a) First, compute the ML estimates for $\mu$ and $\sigma$ for the univariate Gaussian in both the seabass and the salmon case. Also compute the empirical estimates of the priors. (Salmon = 1, Seabass = 2)

**Solution:**

$$L(X_1, X_2, ..., X_N; \mu, \sigma) = (\frac{1}{\sqrt{2\pi}\sigma})^N e^{-\frac{1}{2\sigma^2}\sum_i (X_i - \mu)^2}$$

$$l(X_1, X_2, ..., X_N; \mu, \sigma) = N \ln(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{1}{2\sigma^2}\sum_i (X_i - \mu)^2$$

Solving for $\mu$.

$$\frac{\partial l(\mu, \sigma)}{\partial \mu} = \frac{1}{\sigma^2}(\sum_i X_i - N\mu) = 0$$

$$\hat{\mu} = \frac{1}{N}\sum_i X_i$$

Solving for $\sigma$.

$$\frac{\partial l(\mu, \sigma)}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_i (X_i - \mu)^2 = 0$$

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_i (X_i - \hat{\mu})^2}$$

Plugging in numbers for seabass and salmon: $\mu_1 = 5$, $\mu_2 = 7$, $\sigma_1 = \sqrt{2}$, $\sigma_2 = \sqrt{2}$

Calculating the priors: $\pi_1 = 5/12$, $\pi_2 = 7/12$

$\hat{\mu_1} =$ $\hat{\mu_2} =$

$\hat{\sigma_1} =$ $\hat{\sigma_2} =$

$\hat{\pi_1} =$ $\hat{\pi_2} =$

What is significant about $\hat{\sigma_1}$ and $\hat{\sigma_2}$?

**Solution:** They're the exact same, so a decision boundary between the two Gaussians characterized by them will be linear.

b) Next, find the decision rule when assuming a 0-1 loss function. Recall that a decision rule for the 0-1 loss function will minimize the probability of error.

**Solution:** Recall that assuming a 0-1 loss function results in choosing the class to minimize the probability of error, which means choosing according to this rule:

$$\text{If } \frac{p(w_1|x)}{p(w_2|x)} > 1, \text{ choose } 1$$

Because there is a linear decision boundary, we search for the value such that we classify everything to the right as seabass, and everything to the left as salmon. This boundary is the value of $x$ such that $p(w_1|x) = p(w_2|x)$.

$$p(w_1|x) = p(w_2|x) \implies 5p(x|w_1) = 7p(x|w_2)$$

$$\frac{5}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2}\frac{(x-5)^2}{\sigma^2}\right) = \frac{7}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2}\frac{(x-7)^2}{\sigma^2}\right)$$

$$\ln(5) - \frac{1}{2\sigma^2}(x-5)^2 = \ln(7) - \frac{1}{2\sigma^2}(x-7)^2$$

$$4\ln\left(\frac{5}{7}\right) - x^2 + 10x - 25 = -x^2 + 14x - 49$$

$$4\ln\left(\frac{5}{7}\right) + 24 = 4x$$

$$x = \ln\left(\frac{5}{7}\right) + 6 \approx 5.66$$

The decision rule is: If $x > 5.66$, classify as Seabass! Otherwise classify as Salmon.

Note: Because we had the same variance for both class conditionals, the $x^2$ term canceled out. If that was not the case, then there would be 3 regions, and we would allocate 2 of them to one fish, 1 of them to the other, depending on the height of the posterior probabilities. A good exercise would be to try to draw this: two 1-D Gaussians with different variances.

c) Now, find the decision rule using the loss matrix above. Recall that a decision rule, in general, minimizes the risk, or expected loss.

**Solution:** In the general case, we want to make the decision that minimizes risk. Thus, the decision boundary is located at where the risk of making either decision is equal, or:

$$R(\alpha_1|x) = R(\alpha_2|x)$$

Recall that $R(\alpha_i|x) = \sum_{j=1}^{C} \lambda_{ij} P(w = j|x)$.

$$\lambda_{11} P(w = 1|x) + \lambda_{12} P(w = 2|x) = \lambda_{21} P(w = 1|x) + \lambda_{22} P(w = 2|x)$$

$$2 * P(w = 2|x) = 1 * P(w = 1|x)$$

$$2 * \frac{7}{12} \mathcal{N}(7, 2) = 1 * \frac{7}{12} \mathcal{N}(5, 2)$$

Solving this like part b), we get that $x = 6 + \ln(\frac{5}{14}) \approx 4.97$. Thus, if the weight is greater than 4.97, we classify it as seabass and if not, we classify it as salmon.