Bayesian Decision Theory

Krithika Venkataramani

Salmon and Sea Bass

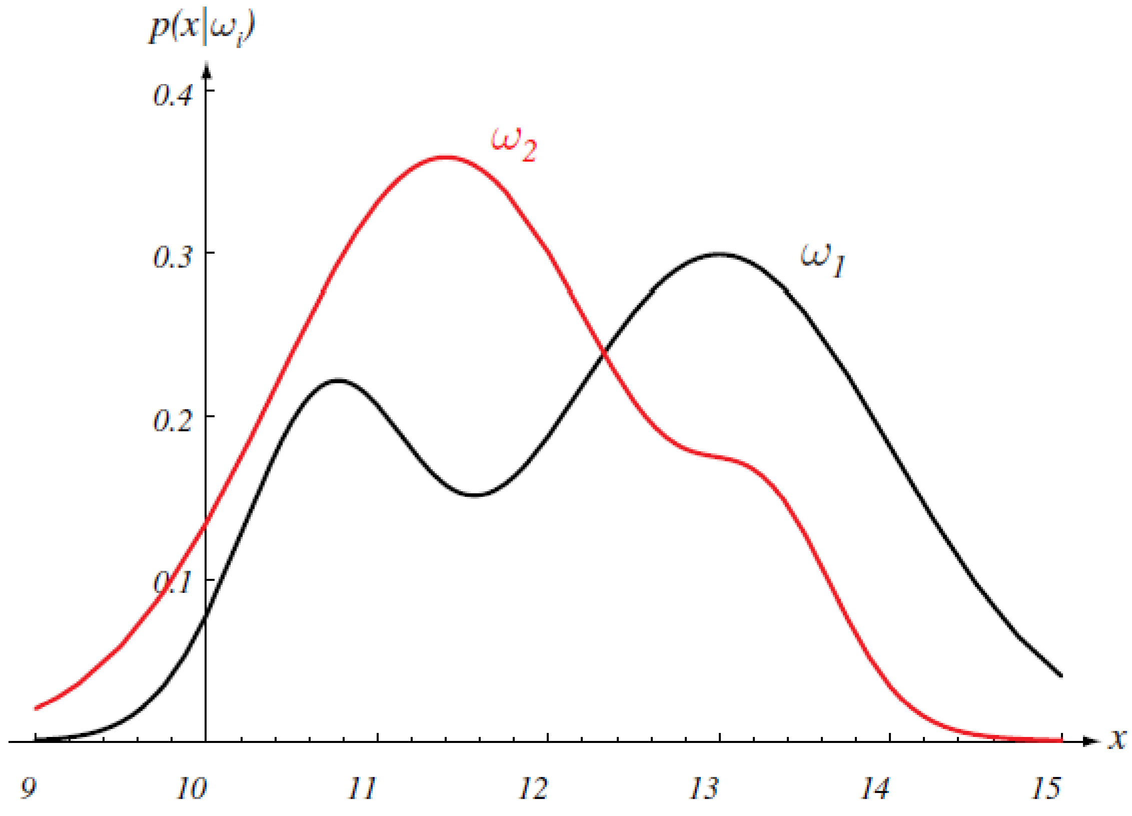
Salmon

Sea Bass

Deciding between Sea Bass and Salmon

- Separate salmon and sea bass on a conveyor present in a fish packing plant
- State of nature: $\omega=\omega_1$ for sea bass and $\omega=\omega_2$ for salmon
- A priori probability: based on the relative numbers in catch
- If there are no other fish, $P(\omega_1)+P(\omega_2)=1$
- Decision without seeing the next fish:
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$
- For one fish, the above decision is OK, but does not seem right for making multiple decisions on all fish
- Additional information: lightness readings
- Class conditional probability density functions: $p(x|\omega_1)$ and $p(x|\omega_2)$

Hypothetical class conditional densities



Class conditional densities on lightness values

Bayes Rule

- Suppose the lightness measurement x is known

$$p(x, \omega_j) = P(\omega_j \mid x) p(x) = p(x \mid \omega_j) P(\omega_j)$$

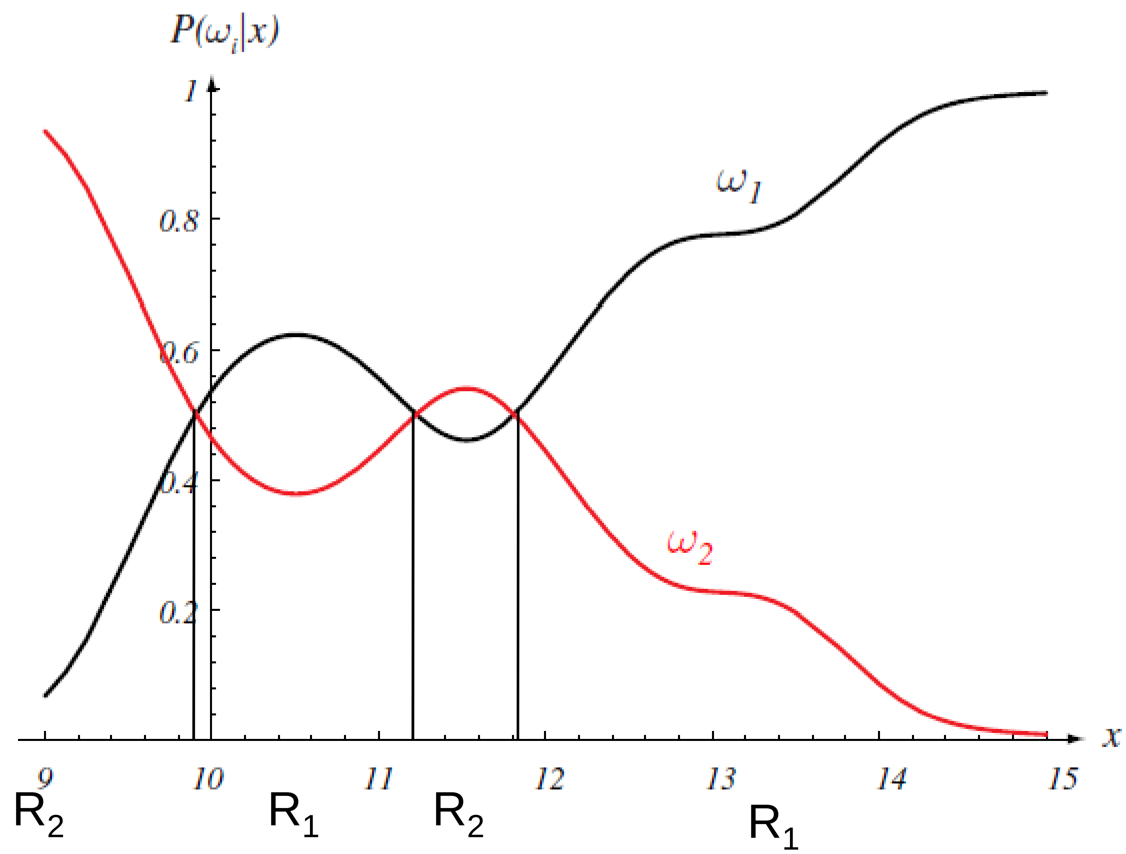
- Bayes Formula: $P(\omega_j \mid x) = \dfrac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$

- For the 2 category case: $p(x) = \sum_{j=1}^{2} p(x \mid \omega_j) P(\omega_j)$

- Bayes formula: $posterior = \dfrac{likelihood \times prior}{evidence}$

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$

Posterior Probabilities used for classification



Posterior probabilities for particular priors $P(\omega_1)=2/3$ and $P(\omega_2)=1/3$

Bayesian Decision Theory Extensions

- Minimum error: Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$
- P(error): $P(\omega_2)$ on deciding ω_1, and $P(\omega_1)$ on deciding ω_2

$$P(error) = \int_{R_1} P(\omega_2 \mid x) p(x) dx + \int_{R_2} P(\omega_1 \mid x) p(x) dx$$

- Extensions
  - More features: d-dimensional feature vector x
  - More categories: c classes - $\omega_1, \omega_2, \ldots, \omega_c$
  - More actions, $\alpha_1, \alpha_2, .., \alpha_a$, other than merely deciding the state of nature
  - Loss functions, $\lambda(\alpha_i|\omega_j)$, more general than error probabilities

Bayes Rule for the general case

- Posterior probability: $P(\omega_j \mid \mathbf{x}) = \dfrac{p(\mathbf{x} \mid \omega_j)P(\omega_j)}{p(\mathbf{x})}$
- Evidence: $p(\mathbf{x}) = \sum_{j=1}^{c} p(\mathbf{x} \mid \omega_j)P(\omega_j)$
- Suppose on observing x, action α_i is taken
- However, the true state of nature could be ω_j
- Conditional risk of taking action α_i on observing x

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j)P(\omega_j \mid \mathbf{x})$$

- Expected Loss/Risk: $R = \int_{\mathbf{x}} R(\alpha(\mathbf{x}) \mid \mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}$
- Minimize expected risk by choosing action α(x) that minimizes conditional risk
- Bayes Decision Rule: Compute conditional risk for all actions and choose the action that minimizes the conditional risk

Two category classification

- Loss of taking action α_i when the true class is ω_j : λ_{ij}
- Conditional risk of taking the two actions

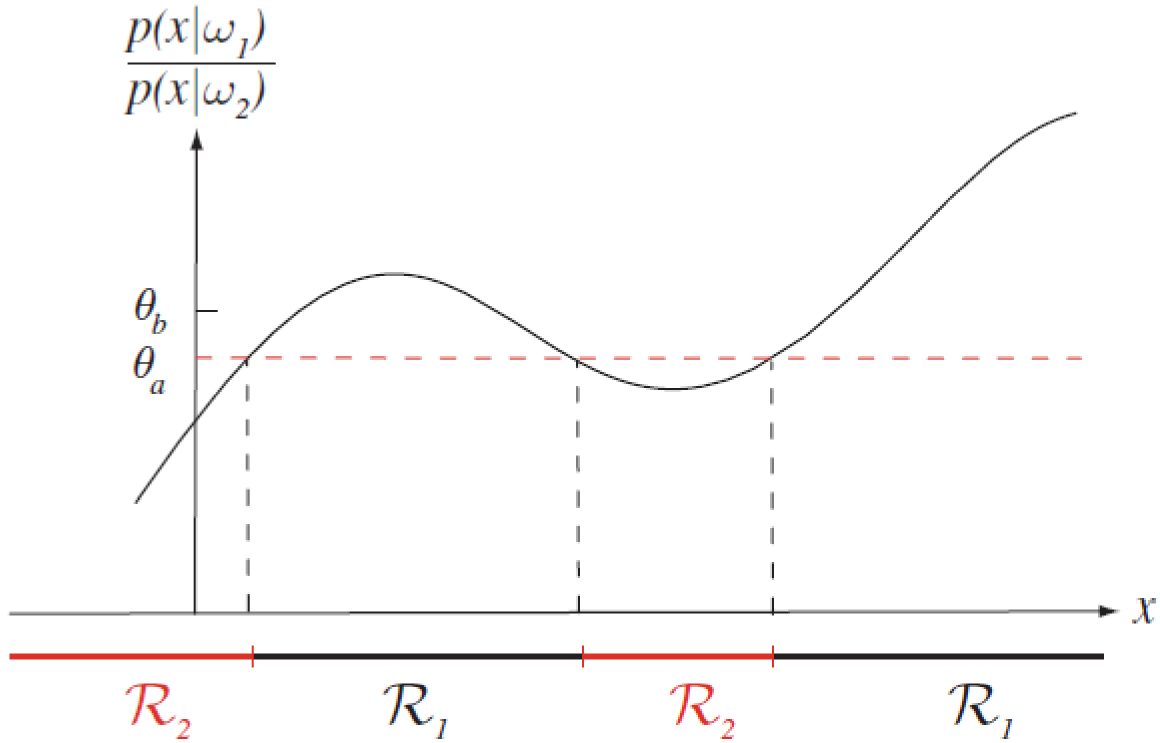
$$R(\alpha_1 \mid \mathbf{x}) = \lambda_{11} P(\omega_1 \mid \mathbf{x}) + \lambda_{12} P(\omega_2 \mid \mathbf{x})$$
$$R(\alpha_2 \mid \mathbf{x}) = \lambda_{21} P(\omega_1 \mid \mathbf{x}) + \lambda_{22} P(\omega_2 \mid \mathbf{x})$$

- Decide ω_1 if $R(\alpha_1|x) < R(\alpha_2|x)$
- In effect, decide ω_1 if

$$(\lambda_{21} - \lambda_{11}) P(\omega_1 \mid \mathbf{x}) > (\lambda_{12} - \lambda_{22}) P(\omega_2 \mid \mathbf{x})$$
$$\text{or, } (\lambda_{21} - \lambda_{11}) p(\mathbf{x} \mid \omega_1) P(\omega_1) > (\lambda_{12} - \lambda_{22}) p(\mathbf{x} \mid \omega_2) P(\omega_2)$$
$$\text{or, } \frac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{P(\omega_2)}{P(\omega_1)}$$

- Likelihood ratio: $\dfrac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)}$

Classification using likelihood ratio



Minimum error rate classification

- Action α_i : decide ω_i
- Decision α_i is correct if true class is ω_i
- Decision α_i is incorrect if true class is ω_j, $j \neq i$
- Loss function of interest: symmetric zero-one loss function

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$$

- All errors are equally costly
- Conditional Risk: $R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x}) = \sum_{j \neq i} P(\omega_j \mid \mathbf{x}) = 1 - P(\omega_i \mid \mathbf{x})$
- For min. error, select α_i that maximizes $P(\omega_i|x)$
- Decide ω_i if $P(\omega_i|x) > P(\omega_j|x)$, $\forall j \neq i$

Discriminant functions

- Assign x to class ω_i if $g_i(x) > g_j(x), \forall j \neq i$
- Discriminant functions: $g_i(x)$
- Examples:
  - $g_i(x) = P(\omega_i|x)$
  - $g_i(x) = p(x|\omega_i)P(\omega_i)$
  - $g_i(x) = \ln p(x|\omega_i) + \ln P(\omega_i)$
  - Two category case: $g(x) = g_1(x) - g_2(x)$

Normal Densities

- Uni-variate Normal:

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

$$E(x) = \int xp(x)dx = \mu$$

$$E[(x-\mu)^2] = \sigma^2$$

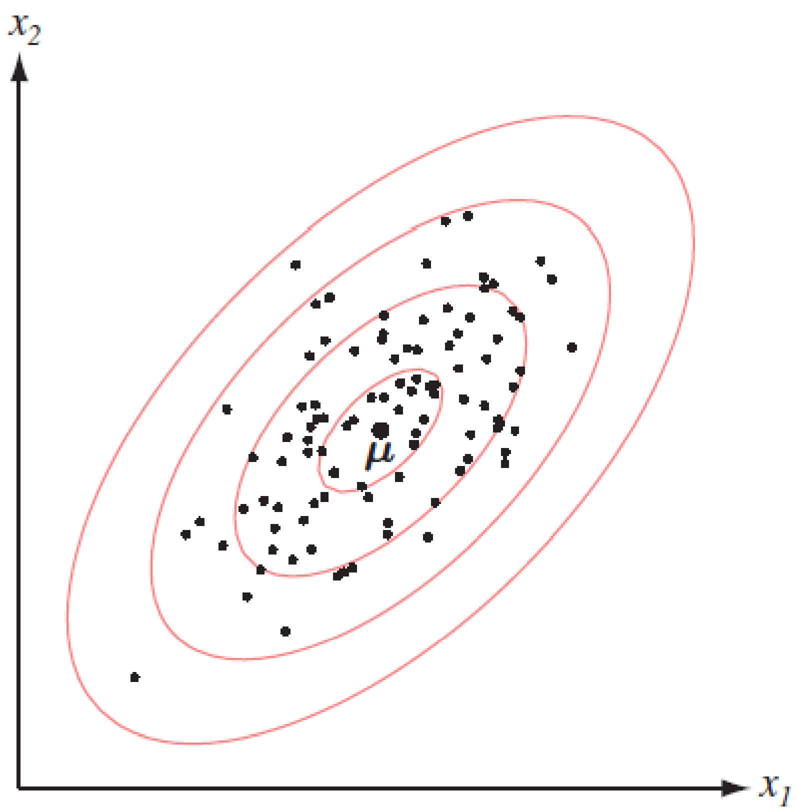
- Multi-variate Normal:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \mid \mathbf{\Sigma} \mid^{1/2}} \exp\left(-\frac{1}{2}\left((\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \right)$$

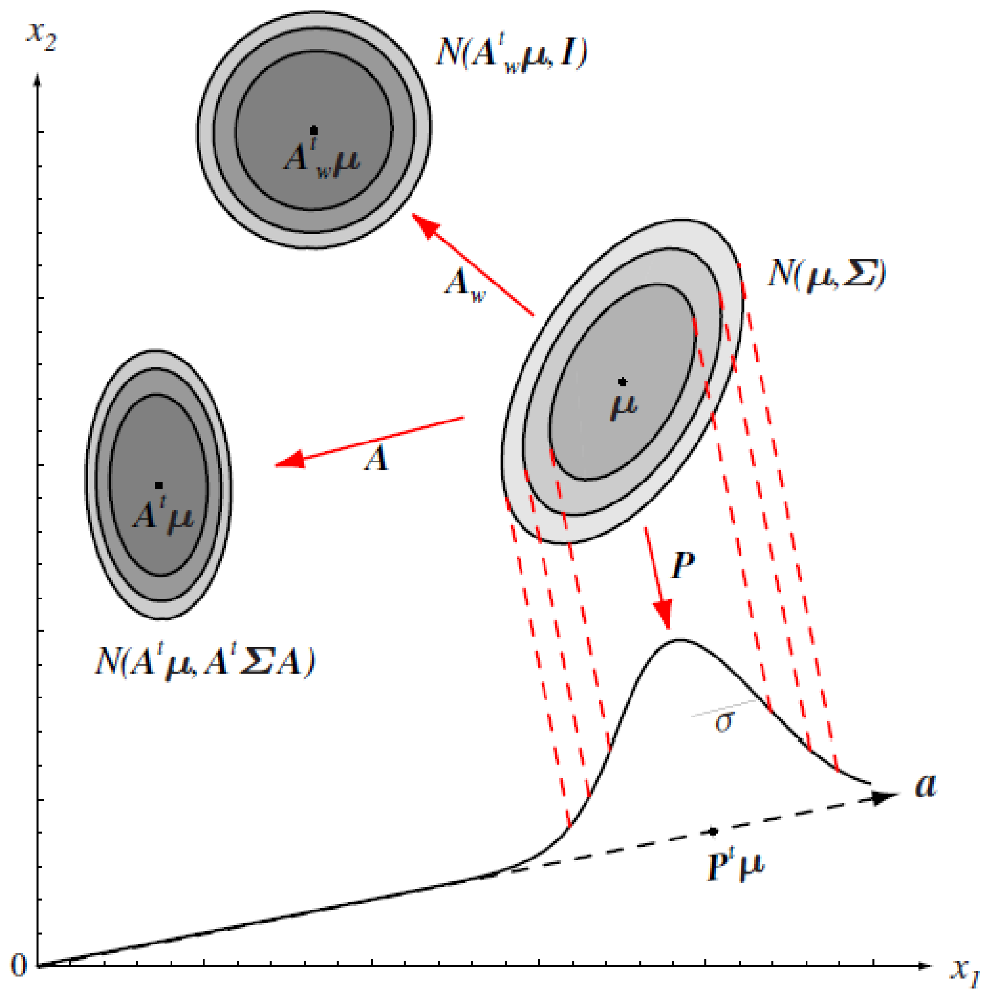
$$\boldsymbol{\mu} = E(\mathbf{x}) = \int \mathbf{x}p(\mathbf{x})d\mathbf{x}$$

$$\mathbf{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] = \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T p(\mathbf{x})d\mathbf{x}$$

Samples from a two-dimensional Gaussian density



Transformations of a Gaussian density



Two dimensional two-category Gaussian densities