1 Decision Theory 201: Class 3

1.1 Today

1. “Minimax” Risk (Loss)
2. Discriminant Functions
3. Categorical Features (Evidence $P(W_i|X)$)
4. Normal (Gaussian Distributions)

1.2 Remember the total expected risk

$$R(\alpha_i|X) = \sum_{j=1}^n \lambda_{ij} \frac{P(X|W_j)P(W_j)}{P(W_j|X)}$$

Consider a plan to choose action $x$ that defines all functions $R_i = (\alpha_i|X)$.

- Overall risk $R = \sum_i R_{\alpha_i} = \sum_i \int_{R_i} \sum_j \lambda_{ij} P(W_j)P(X|W_j)dx$
- In Fish case $R = \int_{R_1} \lambda_{11} P(W_1)P(X|W_1) + \lambda_{12} P(W_2)P(X|W_2)dx + \int_{R_2} \lambda_{21} P(W_1)P(X|W_1)\lambda_{22} P(W_2)P(X|W_2)dx$
- This should get simpler given that:
  $$P(W_1) + P(W_2) = 1$$
  $$\int_{R_1} P(X|W_1)dx + \int_{R_2} P(X|W_2) = 1.$$
- $R = \left\{ \begin{array}{l}
\int_{R_1} \lambda_{11} p_w1 P(X|W_1) + \lambda_{12} (1 - p_w1) P(X|W_2)dx \\
+ \int_{R_2} \lambda_{21} p_w1 P(X|W_1)\lambda_{22} (1 - p_w1) P(X|W_2)dx
\end{array} \right.$$
- $R = \left\{ \begin{array}{l}
\int_{R_1} p_w1 [\lambda_{11} P(X|W_j) - \lambda_{12} P(X|W_2)] + \lambda_{12} P(X|W_2)dx \\
+ \int_{R_2} p_w1 [\lambda_{21} P(X|W_1) - \lambda_{22} P(X|W_2)] + \lambda_{22} P(X|W_2)dx
\end{array} \right.$$
- Using the second relationship:
  $$R = \left\{ \begin{array}{l}
\lambda_{22} + (\lambda_{12} - \lambda_{22}) \int p(X|W_2)dx \\
+ p_w1 [(\lambda_{11} - \lambda_{22}) + \lambda_{21} - \lambda_{11})] \int p(X|W_1)dx \\
- p_w1 [(\lambda_{12} - \lambda_{22}) \int_{R_1} p(X|W_2)dx]
\end{array} \right.$$

Engineering Science 201
Prof: Maurice Smith
1.3 A Drawing

If we want to minimize what the overall risk is. It seems efficient to want to make one term
\[(\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{R^2} P(X|W_1)dx + (\lambda_{12} - \lambda_{22}) \int_{R^2} p(X|W_2)dx = 0\]

We use this as a game for choosing a better prior,best regions.

1.4 Discriminant Functions

\[R(\alpha_i|X) = \sum_j \lambda_{ij} \frac{P(W_j|X)P(W_j)}{P(X)} = \sum_j \lambda_{ij} P(W_j|X)\]

What we coose is to gind a discriminant function \(g_i(X)\) such that for each choice of \(\alpha_1, \alpha_2, \ldots\) is based upon maximizing the discriminant function. For instance, maximize \(g_i(X) = -R(\alpha_i|X)\). As long as discriminant functions have the property that \(g_2(X) > g_1(X)\) under state \(\alpha_2\) then who cares what properties exist in \(g_i(X)\), there need not be a linear mapping frm the Risk function.

1.4.1 One discriminant function

Consider 2-choice \(\lambda_{ii} = 0\) and \(\lambda_{ij} = 1\) if \(i \neq j\) (1 loss if you’re wrong). Use \(g_i(X) = P(X|W_i)P(W_i)\). Then we get:

\[R(\alpha_i|X) = \sum_{j \neq i} \frac{P(X|W_j)P(W_j)}{P(X)}\]

And since \(\frac{P(X)-P(X|W_i)P(W_i)}{P(X)} = 1 - \frac{P(X|W_i)P(W_i)}{P(X)}\) we have something “much nicer” to deal with.

In genera you’re going to have to calculate \(g_1(X), g_2(X), g_3(X)\). Often we want to com-pute just for two categories. The we only need one function:

\[g(X) = \frac{g_1(X)}{g_2(X)}\]

Taking the log function of this will lead us to ask whether a log function is greater or less than zero.

Log Discriminant : \(\log_e P(X|W_1)P(W_1) - \log_e P(X|W_2)P(W_2)\)
1.5 Next time

Next time we’ll talk about categorical features for evidence. Then we’ll get into the Normal Distribution. Since we’ll be talking about multidimensional we’ll have to go to multidimensional discussion.