Bayesian Detection

- Two hypothesis detection problem: $X$ inputs, $D$ classes.
  - $D = -1$, null hypothesis, $f_{X|D}(x|-1)$.
  - $D = 1$, alternate hypothesis, $f_{X|D}(x|1)$.

- Bayesian model
  - Priors known $P(D=-1)=p$, $P(D=1)=1-p$
  - Costs: $C(i,j)$ cost of deciding $Y=i$ given $D=j$

- Objective: partition input space into two sets to minimize average cost
Bayes risk

Proposition: Bayes average risk is minimized by following decision rule

$$f(x) = \text{sign} \ (L(x) - L_0)$$

where $$L(x) = \frac{f_{X|D}(x|1)}{f_{X|D}(x|-1)}$$ is the likelihood ratio and

threshold value given by

$$L_0 = \frac{p/(1-p)}{(C(1,-1)-C(-1,-1))/(C(-1,1)-C(1,1))}$$

Comment: Can also work with log likelihood function

$$l(x) = \log(L(x)).$$
Proof of Proposition

- Let R be the risk then the Bayes average risk is
  \[E(R) = (C(1,1)P(Y=1|D=1) + C(-1,1)P(Y=-1|D=1)P(D=1))
  \[(1-p) + (C(1,-1)P(Y=1|D=-1)P(D=1) - C(-1,-1)P(Y=-1|D=-1)p
  \] = C(-1,1)(1-p) + C(-1,-1)p + \int_{\Omega_1} ((C(1,1)-C(-1,1))(1-p))
  f_{X|D}(x|1) + (C(1,-1) - C(-1,-1))p f_{X|D}(x|-1) dx

- To minimize expected risk add points x to \(\Omega_1\) when term inside integral is negative or \(L(x) \geq L_0\)
Minimum Error Probability

- Costs: $C(-1,-1) = C(1,1) = 0$ and $C(-1,1) = C(1,-1) = 1$.
- Minimum Error probability is equivalent to MAP decision rule
  \[
  \max_i P(D=i|x) = \max_i f_{X|D}(x|i) P(D=i) / f_X(x)
  \]
  from Bayes relationship

- For Gaussian RVs with different means and the same covariance matrix decision rule is given by a linear threshold detector. (Sufficient statistic $t(x) = s^T \Lambda^{-1} x$)
- For Gaussian RVs with different means and different covariance matrix decision rule is given by a quadratic threshold detector.
Bayesian detection comments

- Likelihood ratio $L(x)$ is a sufficient statistic
- Likelihood ratio comparing to threshold $L_0$ minimizes Bayes average risk
- Optimal decision rule is often a linear threshold function (e.g. Gaussian random variables with different means and same covariance functions)
- When you do not have knowledge of prior or likelihood probabilities, then you can learn posterior probability by different approaches
Bayesian Estimation

Given observation $X$ find function $Y(x)$ that approximates $D$ (know $f_{X|D}(x|d)$ likelihood pdf)

- Priors known: $f_D(d)$
- Cost function given: $C(D,Y(X))$
- Minimize expected cost averaged over $X$ and $D$.

$$E(C(D,Y(X))) = \int C(u,Y(x)) f_X(x) f_{D|X}(u|x) \, du \, dx$$

- Estimate $Y(x)$ depends on posterior density $f_{D|X}(d|x)$
- Cost functions: Minimum Mean Squared Error (MMSE), Maximum posterior (MAP), Minimum Absolute Error (MAE)
Minimum Mean Squared Error Estimation

- Minimize $E[(D-Y(X))^2]$ which can be found by minimizing $E[(D-Y(X))^2|X=x]$ for all $x$
- Estimate is the conditional mean estimate given by $Y(X) = E(D|X)$
- Variance of estimate is the expectation of conditional variance given by $E(Var(D|X))$
- Estimate may or may not be a linear function of observations
Bayesian Methods

- Information about knowledge formulated probabilistically
  - Model defined with unknown parameters
  - Specify prior distribution
- Gather data
- Compute posterior distribution
- Use posterior distribution to: make predictions, make decisions, reach conclusions
Finding Posterior Distribution

- Bayes Rule
  \[ P(\text{param.}|\text{data}) = P(\text{data}|\text{param.})P(\text{param.})/P(\text{data}) \]
- Posterior \( \propto \) Likelihood \( \times \) Prior
- To make predictions on new data
  \[ P(\text{new data}|\text{data}) = \int P(\text{new data}|\text{param.})P(\text{param.}|\text{data}) \]
Representing Priors and Posterior Dist.

- Priors and posterior distributions often have complex distributions that are not easily represented
- Represent distributions using samples
  - Obtaining a sample from priors
  - Obtaining a sample from posterior distribution (more difficult)
- Example: A Hard Linear Classifier (Radford Neal, NIPS 2004 tutorial)