EE269
Signal Processing for Machine Learning

Lecture 6

Instructor : Mert Pilanci

Stanford University

Jan 9 2019

Bayes classifiers

Given P_{xy} joint probability distribution of $(x[n], y)$ what is the *best classifier* ?

- signal classifier function $f(x[n]) \quad : \quad \mathbb{R}^N \rightarrow \{1, ..., k\}$
- best classifier depends on the performance measure
- risk (probability of error):

$$R(f) = P_{xy}(f(x) \neq y)$$

Bayes risk R^*

- **Bayes risk** is R^* the smallest risk of any classifier
  If $R(f) = R^*$, then f is a **Bayes classifier**

Bayes risk R^*

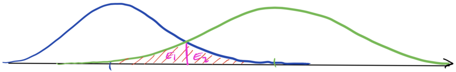
- **Bayes risk** is R^* the smallest risk of any classifier
  If $R(f) = R^*$, then f is a **Bayes classifier**

  $\pi_k := P_y(y = k)$ prior class probabilities
  $g_k(x) := P_{x|y=k}$ class conditional distribution
- define **posterior** class probabilities:

$$\eta_k(x) = P_{y=k|x}(y = k|x)$$

note that $\sum_{k=1}^{K} \eta_k(x) = 1$

Example: $x[n] = x \in \mathbb{R}$



- Suppose $x|y$ is Gaussian, $y \in \{-1, +1\}$

$$x|y = {\scriptstyle -1} \sim N(\mu_-, \sigma_-^2) \text{ and } x|y = {\scriptstyle +1} \sim N(\mu_+, \sigma_+^2)$$

$$\eta_1(x) = P(y = 1|x) = \frac{P(x|y{=}1)P(y{=}1)}{P(x)} \text{ (Bayes rule)}$$

Bayes classifier

$\pi_k := P_y(y = k)$ prior class probabilities

$g_k(x) := P_{x|y=k}$ class conditional distribution

$\eta_k(x) = P_{y=k|x}(y = k|x)$ posterior class probabilities

Theorem: The classifier

$$\begin{aligned} f^*(x) &= \arg \max_{k=1,\ldots,K} \eta_k(x) \\ &= \arg \max_{k=1,\ldots,K} \pi_k g_k(x) \end{aligned}$$

is a Bayes classifier

Detecting a constant signal in Gaussian noise

- $x[n] = [x_1, ...x_N] \sim N(0, \sigma^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [x_1, ...x_N] \sim N(\mu, \sigma^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume μ is known

Detecting a constant signal in Gaussian noise

- $x[n] = [x_1, ...x_N] \sim N(0, \sigma^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [x_1, ...x_N] \sim N(\mu, \sigma^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume μ is known

  $g_1(x) := P_{x|y=1} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i)^2}$

  $g_2(x) := P_{x|y=2} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$

Detecting a constant signal in Gaussian noise

- $x[n] = [x_1, ...x_N] \sim N(0, \sigma^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [x_1, ...x_N] \sim N(\mu, \sigma^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume μ is known

  $g_1(x) := P_{x|y=1} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i)^2}$

  $g_2(x) := P_{x|y=2} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$

  Bayes classifier:

$$\text{Classify as signal+noise if} \quad \pi_1 g_1(x) < \pi_2 g_2(x)$$

$$\mu \sum_{i=1}^{N} x_i > \sigma^2 \log(\frac{\pi_1}{\pi_2}) + N\mu^2/2$$

Detecting a change in variance

- $x[n] = [x_1, ...x_N] \sim N(0, \sigma_1^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [x_1, ...x_N] \sim N(0, \sigma_2^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume σ_1 and σ_2 is known

Detecting a change in variance

- $x[n] = [x_1, ...x_N] \sim N(0, \sigma_1^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [x_1, ...x_N] \sim N(0, \sigma_2^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume σ_1 and σ_2 is known

  $g_1(x) := P_{x|y=1} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i)^2}$

  $g_2(x) := P_{x|y=2} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x_i)^2}$

Detecting a change in variance

- $x[n] = [x_1, ...x_N] \sim N(0, \sigma_1^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [x_1, ...x_N] \sim N(0, \sigma_2^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume σ_1 and σ_2 is known

  $g_1(x) := P_{x|y=1} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i)^2}$

  $g_2(x) := P_{x|y=2} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(x_i)^2}$

  Bayes classifier:

$$\text{Classify as class 2 if } \pi_1 g_1(x) < \pi_2 g_2(x)$$

$$\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right) \sum_{i=1}^{N} x_i^2 > \log(\frac{\pi_1}{\pi_2}) - \frac{N}{2} \log(\frac{\sigma_1^2}{\sigma_2^2})$$

Detecting a known waveform

- $s[n] \in \mathbb{R}^N$ known waveform
- $x[n] = [x_1, ...x_N] \sim N(0, \sigma^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [s_1, ...s_N] + N(0, \sigma^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume μ is known

Detecting a known waveform

- $s[n] \in \mathbb{R}^N$ known waveform
- $x[n] = [x_1, ...x_N] \sim N(0, \sigma^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [s_1, ...s_N] + N(0, \sigma^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume μ is known

  $g_1(x) := P_{x|y=1} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i)^2}$

  $g_2(x) := P_{x|y=2} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - s_i)^2}$

Detecting a known waveform

- $s[n] \in \mathbb{R}^N$ known waveform
- $x[n] = [x_1, ... x_N] \sim N(0, \sigma^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [s_1, ... s_N] + N(0, \sigma^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume μ is known

  $g_1(x) := P_{x|y=1} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i)^2}$

  $g_2(x) := P_{x|y=2} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - s_i)^2}$

  Bayes classifier:

$$\text{Classify as signal+noise if} \quad \pi_1 g_1(x) < \pi_2 g_2(x)$$

$$\sum_{i=1}^N x_i s_i > \sigma^2 \log(\frac{\pi_1}{\pi_2}) + \frac{1}{2} \sum_{i=1}^N |s_i|^2$$

Detecting a known waveform

- $s[n] \in \mathbb{R}^N$ known waveform
- $x[n] = [x_1, ...x_N] \sim N(0, \sigma^2)$ i.i.d. for $y = 1$ (noise)
- $x[n] = [s_1, ...s_N] + N(0, \sigma^2)$ i.i.d. for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume μ is known

  $g_1(x) := P_{x|y=1} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i)^2}$

  $g_2(x) := P_{x|y=2} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - s_i)^2}$

  Bayes classifier:

$$\text{Classify as signal+noise if} \quad \pi_1 g_1(x) < \pi_2 g_2(x)$$

$$\sum_{i=1}^{N} x_i s_i > \sigma^2 \log(\frac{\pi_1}{\pi_2}) + \frac{1}{2} \sum_{i=1}^{N} |s_i|^2$$

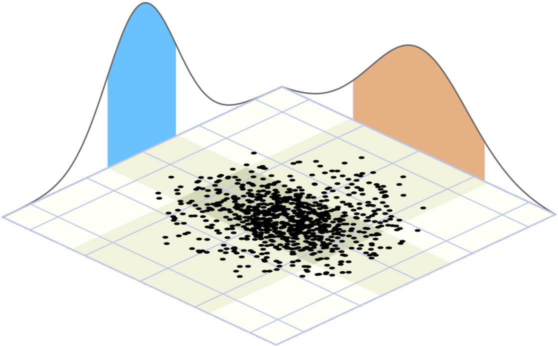
$$\langle x, s \rangle > \sigma^2 \log(\frac{\pi_1}{\pi_2}) + \frac{1}{2} ||s||_2^2$$

- **matched filter** (e.g., face detector, DFT basis)

Multivariate Gaussian

$$x_1, ...x_n \sim \text{ i.i.d. } N(\mu, \Sigma)$$

$$P(x_1, ...x_n; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$



Detecting a correlated signal

- $[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
- $[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume σ and Σ are known

Detecting a correlated signal

- $[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
- $[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume σ and Σ are known

  $$g_1(x) := P_{x|y=1} = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2}x^T(\sigma^2 I_n)^{-1}x}$$

  $$g_2(x) := P_{x|y=2} = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}x^T(\Sigma+\sigma^2 I_n)^{-1}x}$$

Detecting a correlated signal

- $[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
- $[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)
- $P(y = 1) = \pi_1$ and $P(y = 2) = \pi_2$

  Assume σ and Σ are known

  $g_1(x) := P_{x|y=1} = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2}x^T(\sigma^2 I_n)^{-1}x}$

  $g_2(x) := P_{x|y=2} = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}x^T(\Sigma+\sigma^2 I_n)^{-1}x}$

  Bayes classifier:

$$\text{Classify as signal+noise if} \quad \pi_1 g_1(x) < \pi_2 g_2(x)$$

$$\log(\pi_1) - \frac{1}{2}x(\sigma^2 I_n)^{-1}x < -\frac{1}{2}x(\Sigma+\sigma^2 I_n)^{-1}x + \log\pi_2 + \text{constant}$$

Detecting a correlated signal

- $[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
- $[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)

$$\log(\pi_1) - \frac{1}{2} x^T (\sigma^2 I_n)^{-1} x < -\frac{1}{2} x^T (\Sigma + \sigma^2 I_n)^{-1} x + \log \pi_2 + \text{constant}$$

- Matrix inversion lemma (A invertible)
  $(A + B)^{-1} = A^{-1} - A^{-1}(I + BA^{-1})^{-1} BA^{-1}$

Detecting a correlated signal

- $[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
- $[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)

$$\log(\pi_1) - \frac{1}{2} x^T (\sigma^2 I_n)^{-1} x < -\frac{1}{2} x^T (\Sigma + \sigma^2 I_n)^{-1} x + \log \pi_2 + \text{constant}$$

- Matrix inversion lemma (A invertible)
  $(A + B)^{-1} = A^{-1} - A^{-1}(I + BA^{-1})^{-1} BA^{-1}$
- $(\sigma^2 I_n)^{-1} - (\Sigma + \sigma^2 I_n)^{-1} = \frac{1}{\sigma^2}(\sigma^2 I_n + \Sigma)^{-1} \Sigma$

Detecting a correlated signal

- $[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
- $[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)

$$\log(\pi_1) - \frac{1}{2} x^T (\sigma^2 I_n)^{-1} x < -\frac{1}{2} x^T (\Sigma + \sigma^2 I_n)^{-1} x + \log \pi_2 + \text{constant}$$

- Matrix inversion lemma (A invertible)
  $(A + B)^{-1} = A^{-1} - A^{-1}(I + BA^{-1})^{-1} BA^{-1}$
- $(\sigma^2 I_n)^{-1} - (\Sigma + \sigma^2 I_n)^{-1} = \frac{1}{\sigma^2} (\sigma^2 I_n + \Sigma)^{-1} \Sigma$
- Classify as signal when
  $x^T (\sigma^2 I_n + \Sigma)^{-1} \Sigma x > \sigma^2 \log \frac{\pi_1}{\pi_2} + \text{constant}$

Detecting a correlated signal

$[x_1, ... x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)

$[x_1, ... x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)

Classify as signal when

$x^T(\sigma^2 I_n + \Sigma)^{-1} \Sigma x > \sigma^2 \log \frac{\pi_1}{\pi_2} + \text{constant}$

- Eigenvalue Decomposition: $\Sigma = U \Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$

Detecting a correlated signal

$[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
$[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)
Classify as signal when
$x^T(\sigma^2 I_n + \Sigma)^{-1}\Sigma x > \sigma^2 \log \frac{\pi_1}{\pi_2} + \text{constant}$

- Eigenvalue Decomposition: $\Sigma = U\Lambda U^T = \sum_{i=1}^n \lambda_i u_i u_i^T$

$$x^T(\sigma^2 I_n + U\Lambda U^T)^{-1} U\Lambda U^T x$$
$$= x^T U(\sigma^2 I_n + \Lambda)^{-1}\Lambda U^T x$$

Detecting a correlated signal

$[x_1, ...x_N] \sim N(0, \sigma^2 I_n)$ for $y = 1$ (noise)
$[x_1, ...x_N] \sim N(0, \Sigma) + N(0, \sigma^2 I_n)$ for $y = 2$ (signal+noise)
Classify as signal when
$x^T(\sigma^2 I_n + \Sigma)^{-1}\Sigma x > \sigma^2 \log \frac{\pi_1}{\pi_2} + \text{constant}$

- Eigenvalue Decomposition: $\Sigma = U\Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$

$$x^T(\sigma^2 I_n + U\Lambda U^T)^{-1}U\Lambda U^T x$$
$$= x^T U(\sigma^2 I_n + \Lambda)^{-1}\Lambda U^T x$$

- Change of basis: $y = U^T x$
- $y^T(\sigma^2 I_n + \Lambda)^{-1}\Lambda y = \sum_{n=0}^{N-1} \frac{\lambda_n}{\lambda_n + \sigma^2}|y[n]|^2$

$y_{\text{filtered}}[n] = \sqrt{\frac{\lambda_n}{\lambda_n + \sigma^2}} y[n]$

Optimality of filtering for circulant covariances

$$x^T U(\sigma^2 I_n + \Lambda)^{-1} \Lambda U^T x > \sigma^2 \log \frac{\pi_1}{\pi_2}$$

- Symmetric circulant covariances

$$\Sigma = \begin{bmatrix} s[0] & s[1] & s[2] & \dots & s[2] & s[1] \\ s[1] & s[0] & s[1] & \dots & s[3] & s[2] \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s[2] & s[3] & s[4] & \dots & s[0] & s[1] \\ s[1] & s[2] & s[3] & \dots & s[1] & s[0] \end{bmatrix}$$

- $\Sigma_{ij} = \Sigma_{ji}$
- $s[N-n] = s[n]$

Optimality of filtering for circulant covariances

$$x^T U(\sigma^2 I_n + \Lambda)^{-1} \Lambda U^T x > \sigma^2 \log \frac{\pi_1}{\pi_2}$$

- Symmetric circulant covariances

$$\Sigma = \begin{bmatrix} s[0] & s[1] & s[2] & \dots & s[2] & s[1] \\ s[1] & s[0] & s[1] & \dots & s[3] & s[2] \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s[2] & s[3] & s[4] & \dots & s[0] & s[1] \\ s[1] & s[2] & s[3] & \dots & s[1] & s[0] \end{bmatrix}$$

- $\Sigma_{ij} = \Sigma_{ji}$
- $s[N-n] = s[n]$

$$\text{Fourier basis: } w_k[n] = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}nk}$$

$$w_k^H \Sigma = S[k] w_k^H$$