EE269
Signal Processing for Machine Learning

Lecture 7

Instructor : Mert Pilanci

Stanford University

January 30, 2019

Stationary Random Signals

- Mean: $m_x[n] = \mathbb{E}\big[x[n]\big]$

Stationary Random Signals

- Mean: $m_x[n] = \mathbb{E}\big[x[n]\big]$
- Covariance (autocovariance):
  $\Sigma_x[k,l] = \mathbb{E}\big[(x[k] - m_x[k])(x[l] - m_x[l])^*\big]$

Stationary Random Signals

- Mean: $m_x[n] = \mathbb{E}\big[x[n]\big]$
- Covariance (autocovariance):
  $\Sigma_x[k,l] = \mathbb{E}\big[(x[k] - m_x[k])(x[l] - m_x[l])^*\big]$
- Assume $m_x[n] = 0$
- $\Sigma_x[k,l] = \mathbb{E}\big[x[k]x^*[l]\big]$

Stationary Random Signals

- Mean: $m_x[n] = \mathbb{E}\big[x[n]\big]$
- Covariance (autocovariance):
  $\Sigma_x[k,l] = \mathbb{E}\big[(x[k] - m_x[k])(x[l] - m_x[l])^*\big]$
- Assume $m_x[n] = 0$
- $\Sigma_x[k,l] = \mathbb{E}\big[x[k]x^*[l]\big]$
- **Wide Sense Stationarity**:

  $\Sigma_x[k,l]$ depends on the difference $k - l$

Stationary Random Signals

- Mean: $m_x[n] = \mathbb{E}\big[x[n]\big]$
- Covariance (autocovariance):
  $\Sigma_x[k,l] = \mathbb{E}\big[(x[k] - m_x[k])(x[l] - m_x[l])^*\big]$
- Assume $m_x[n] = 0$
- $\Sigma_x[k,l] = \mathbb{E}\big[x[k]x^*[l]\big]$
- **Wide Sense Stationarity**:

  $\Sigma_x[k,l]$ depends on the difference $k - l$
- Autocorrelation: $r_x[k-l] = \mathbb{E}\big[x[k]x^*[l]\big]$

Example random signal

- $x[n] = A\sin(nw_0 + \phi)$
  $A \sim N(0, \sigma^2)$, and ϕ uniform over the interval $[-\pi, \pi]$
- $\mathbb{E}\big[x[n]\big] = \mathbb{E}_A\big[A\big]\mathbb{E}_\phi\big[\sin(nw_0 + \phi)\big]$

Example random signal

- $x[n] = A\sin(nw_0 + \phi)$
  $A \sim N(0, \sigma^2)$, and ϕ uniform over the interval $[-\pi, \pi]$
- $\mathbb{E}\big[x[n]\big] = \mathbb{E}_A\big[A\big]\mathbb{E}_\phi\big[\sin(nw_0 + \phi)\big]$

$$\begin{aligned}
\Sigma_x[k, l] &= \mathbb{E}\big[x[k]x^*[l]\big] \\
&= \mathbb{E}\big[A\sin(kw_0 + \phi)A\sin(lw_0 + \phi)\big] \\
&= \mathbb{E}_A\big[A^2\big]\mathbb{E}_\phi\big[\sin(kw_0 + \phi)\sin(lw_0 + \phi)\big]
\end{aligned}$$

Example random signal

- $x[n] = A\sin(nw_0 + \phi)$
  $A \sim N(0, \sigma^2)$, and ϕ uniform over the interval $[-\pi, \pi]$
- $\mathbb{E}\big[x[n]\big] = \mathbb{E}_A\big[A\big]\mathbb{E}_\phi\big[\sin(nw_0 + \phi)\big]$

$$\begin{aligned}\Sigma_x[k, l] &= \mathbb{E}\big[x[k]x^*[l]\big] \\ &= \mathbb{E}\big[A\sin(kw_0 + \phi)A\sin(lw_0 + \phi)\big] \\ &= \mathbb{E}_A\big[A^2\big]\mathbb{E}_\phi\big[\sin(kw_0 + \phi)\sin(lw_0 + \phi)\big]\end{aligned}$$

- Trigonometric identity:
  $2\sin\alpha\sin\beta = \cos(\alpha - \beta) - \cos(\alpha + \beta)$

$$\begin{aligned}\Sigma_x[k, l] &= \mathbb{E}\big[\sin(kw_0 + \phi)\sin(lw_0 + \phi)\big] \\ &= \frac{\sigma^2}{2}\mathbb{E}\big[\cos((k - l)w_0)\big] - \frac{\sigma^2}{2}\mathbb{E}\big[\cos((k + l)w_0 + 2\phi)\big] \\ &= \frac{\sigma^2}{2}\big[\cos((k - l)w_0)\big]\end{aligned}$$

- $x[n] = A\sin(nw_0 + \phi)$
- $r[k-l] = \frac{\sigma^2}{2}\left[\cos((k-l)w_0)\right]$
- Covariance matrix $\Sigma[k,l]$ is Toeplitz for stationary signals

$$\Sigma = \begin{bmatrix} r[0] & r^*[1] & \dots & r^*[N-1] & \\ r[1] & r[0] & \dots & r^*[N-2] & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r[2] & r[N-3] & \dots & r^*[1] & \\ r[N-1] & r[N-2] & \dots & r^*[0] & \end{bmatrix}$$

- Circulant if $r[N-k] = r^*[k]$
- $w_0 = \frac{2\pi}{N} \times \text{integer}$

Linear and Quadratic Discriminant Analysis

- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma)$ when $y = k$
  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k)}$
- K classes

$$f(x) = \arg \max_{k=1,\ldots,K} \pi_k g_k(x)$$

Linear Discriminant Analysis

- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma)$ when $y = k$

  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k)}$
- Two classes: Classify as class 1 if

$$\log \pi_1 - \frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1) > \log \pi_2 - \frac{1}{2}(x-\mu_2)^T \Sigma^{-1} (x-\mu_2)$$

Linear Discriminant Analysis

- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma)$ when $y = k$
  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k)}$
- Two classes: Classify as class 1 if

$$\log \pi_1 - \frac{1}{2}(x-\mu_1)^T \Sigma^{-1} (x-\mu_1) > \log \pi_2 - \frac{1}{2}(x-\mu_2)^T \Sigma^{-1} (x-\mu_2)$$

- $\Sigma = I$, $\pi_1 = \pi_2 = \frac{1}{2}$

$$-\frac{1}{2}(x-\mu_1)^T (x-\mu_1) > -\frac{1}{2}(x-\mu_2)^T (x-\mu_2)$$

Scaled identity covariances $\Sigma_k = \sigma^2 I$

- Suppose $x[n] = [x_1, ... x_N] \sim N(\mu_k, \Sigma_k)$ when $y = k$

  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}$
- K classes
- $h_k(x) = w_k^T x + w_{k0}$
- Classify as class k if $h_k(x) > h_{k'}(x) \quad \forall k' \neq k$

  $w_k = \frac{1}{\sigma^2} \mu_k$

  $w_{k0} = -\frac{1}{2\sigma^2} \mu_k^T \mu_k + \log \pi_k$

Scaled identity covariances $\Sigma_k = \sigma^2 I$

- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma_k)$ when $y = k$
  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}$
- K classes
- Decision boundary: hyperplane

$$w^T(x - x_0) = 0$$

$$w = \mu_i - \mu_j$$
$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{||\mu_i - \mu_j||^2} \log \frac{\pi_i}{\pi_j}(\mu_i - \mu_j)$$

Scaled identity covariances $\Sigma_k = \sigma^2 I$

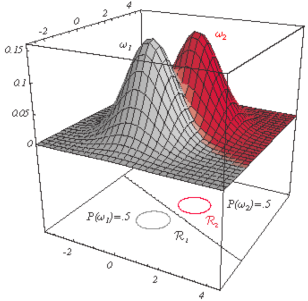
- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma_k)$ when $y = k$
  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}$
- K classes
- Decision boundary: hyperplane

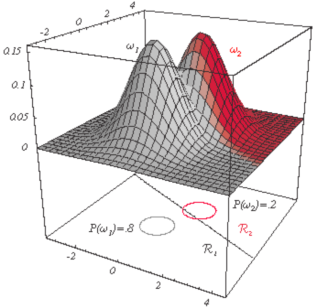
$$w^T(x - x_0) = 0$$

$$w = \mu_i - \mu_j$$
$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{||\mu_i - \mu_j||^2} \log \frac{\pi_i}{\pi_j}(\mu_i - \mu_j)$$

- Hyperplane passes through the point x_0 and is orthogonal to w

Scaled identity covariances $\Sigma_k = \sigma^2 I$





Scaled identity covariances $\Sigma_k = \sigma^2 I$

- If $\pi_k = \frac{1}{K}$, equivalent to minimum distance classifier

$$\text{class } k^* \text{ where } k^* = \arg\min_k ||x - \mu_k||_2$$

Identical covariances $\Sigma_k = \Sigma$

- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma)$ when $y = k$

  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)}$
- $h_k(x) = w_k^T x + w_{k0}$
- Classify as class k if $h_k(x) > h_{k'}(x) \quad \forall k' \neq k$

  $w_k = \Sigma^{-1}\mu_k$

  $w_{k0} = -\frac{1}{2}\mu_k^T \Sigma^{-1} \mu_k^T + \log \pi_k$

Identical covariances $\Sigma_k = \Sigma$

- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma)$ when $y = k$
  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)}$
- Decision boundary: hyperplane

$$w^T(x - x_0) = 0$$

$$w = \Sigma^{-1}(\mu_i - \mu_j)$$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\log \frac{\pi_i}{\pi_j}}{(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$$

Identical covariances $\Sigma_k = \Sigma$

- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma)$ when $y = k$
  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)}$
- Decision boundary: hyperplane

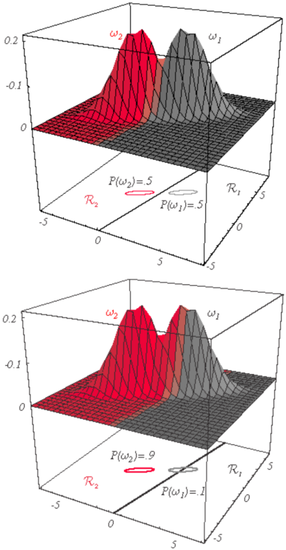
$$w^T(x - x_0) = 0$$

$$w = \Sigma^{-1}(\mu_i - \mu_j)$$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\log \frac{\pi_i}{\pi_j}}{(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$$

- Hyperplane passes through x_0 but not necessarily orthogonal to the lines between the means

Identical covariances $\Sigma_k = \Sigma$



Mahalanobis distance

$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

Δ = Mahalanobis distance from μ to x

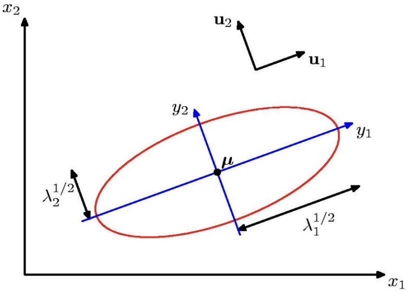
$$\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^{\mathrm{T}}$$

where $(\mathbf{u}_i, \lambda_i)$ are the ith eigenvector and eigenvalue of Σ.

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

$$y_i = \mathbf{u}_i^{\mathrm{T}} (\mathbf{x} - \boldsymbol{\mu})$$

or $\mathbf{y} = \mathbf{U}(\mathbf{x} - \mu)$



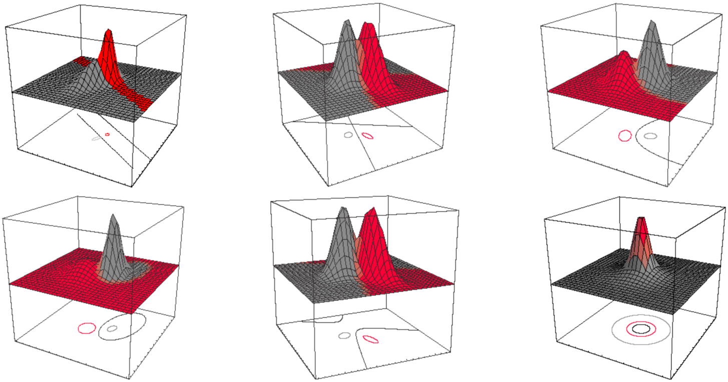
Whitening

- $\Sigma = U\Lambda^{-1}U$
- $\Sigma^{-\frac{1}{2}} = U\Lambda^{-\frac{1}{2}}U$
- Mahalanobis distance
- $d_M(x, y) = ||\Sigma^{-\frac{1}{2}}(x - y)||_2$

Quadratic Discriminant Analysis: Σ_k arbitrary

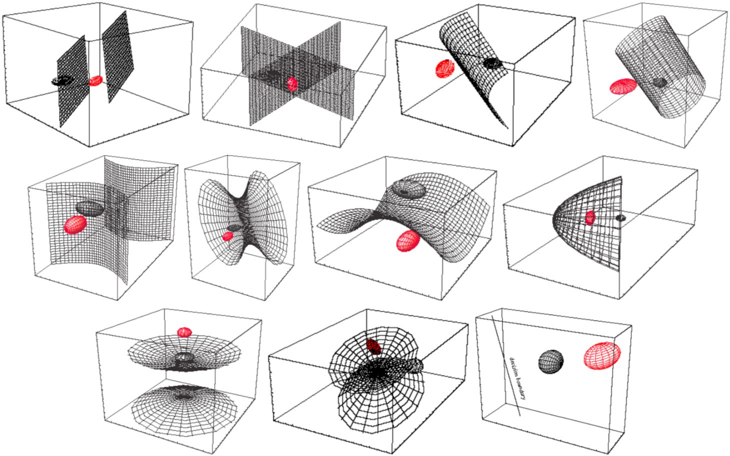
- Suppose $x[n] = [x_1, ...x_N] \sim N(\mu_k, \Sigma_k)$ when $y = k$
  $g_k(x) = P_{x|y=k} = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}$
- $h_k(x) = x^T W_k x + w_k^T x + w_{k0}$
- Classify as class k if $h_k(x) > h_{k'}(x) \quad \forall k' \neq k$
  $W_k = -\frac{1}{2}\Sigma_k^{-1}$
  $w_k = \Sigma_k^{-1}\mu_k$
  $w_{k0} = -\frac{1}{2}\mu_k^T \Sigma_k^{-1} \mu_k - \frac{1}{2}\log|\Sigma_k| + \log \pi_k$

Quadratic decision regions: hyperquadrics



slide credit: S. Aksoy

Quadratic decision regions: hyperquadrics



slide credit: S. Aksoy

Estimating parameters: univariate Gaussian

- Suppose $x_1, x_2, ...x_n$ i.i.d. $\sim N(\mu, \sigma^2)$
- Estimating means

  $\mu_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_n$
- Estimating variances

  $\sigma^2_{ML} = \frac{1}{n} \sum_{i=1}^{n} (x_n - \mu_{ML})^2$

Estimating parameters: multivariate Gaussian

- Suppose $x_1, x_2, ...x_n$ i.i.d. $\sim N(\mu, \Sigma)$
- Estimating means

  $\mu_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_n$

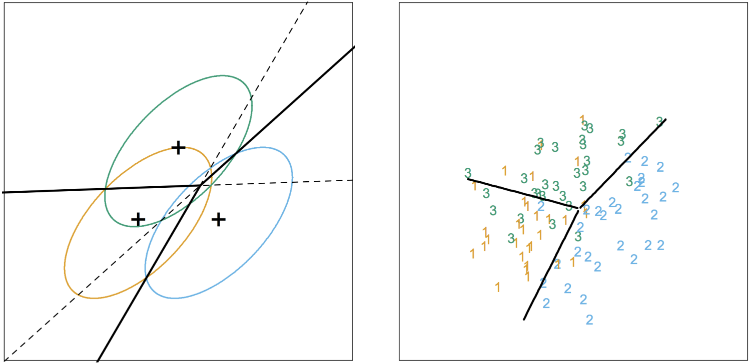
Estimating parameters: multivariate Gaussian

- Suppose $x_1, x_2, ...x_n$ i.i.d. $\sim N(\mu, \Sigma)$
- Estimating means

  $\mu_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_n$
- Estimating covariances

  $\Sigma_{ML} = \frac{1}{n} \sum_{i=1}^{n} (x_n - \mu_{ML})(x_n - \mu_{ML})^T$

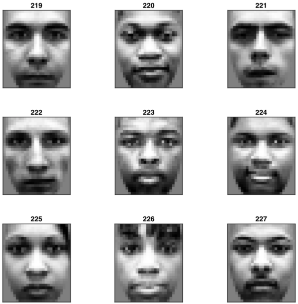
Examples



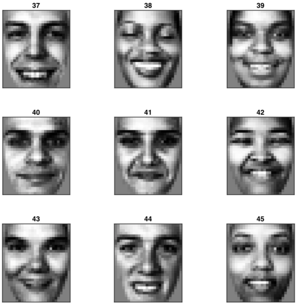
slide credit: T. Hastie et al.

Examples

Some non-smiling samples

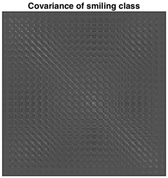


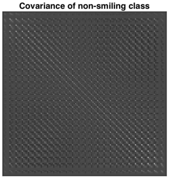
Some smiling samples



slide credit: P. Smaragdis

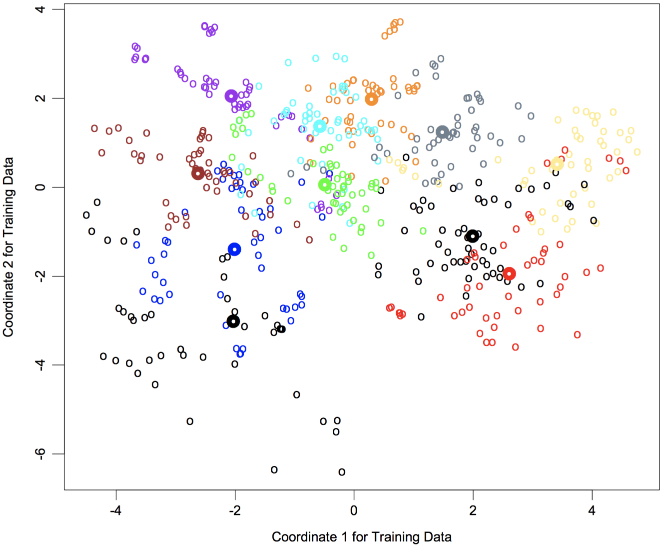
Examples





slide credit: P. Smaragdis

Examples



slide credit: T. Hastie et al.

Examples

Technique	Error Rates	
	Training	Test
Linear regression	0.48	0.67
Linear discriminant analysis	0.32	0.56
Quadratic discriminant analysis	0.01	0.53

Linear vs Quadratic Discriminant Analysis

- LDA

  Estimate μ_k, for $k = 1..., K$ and Σ

  $Kn + \binom{n}{2} + n$ parameters

Linear vs Quadratic Discriminant Analysis

- LDA

  Estimate μ_k, for $k = 1..., K$ and Σ

  $Kn + \binom{n}{2} + n$ parameters

- QDA

  Estimate μ_k, $\Sigma_k \quad$ for $k = 1..., K$

  $Kn + K\left(\binom{n}{2} + n\right)$ parameters