Prof. Marios Savvides

Pattern Recognition Theory

Lecture 3: Decision Theory III

Electrical & Computer ENGINEERING

Minimum Error Classifier

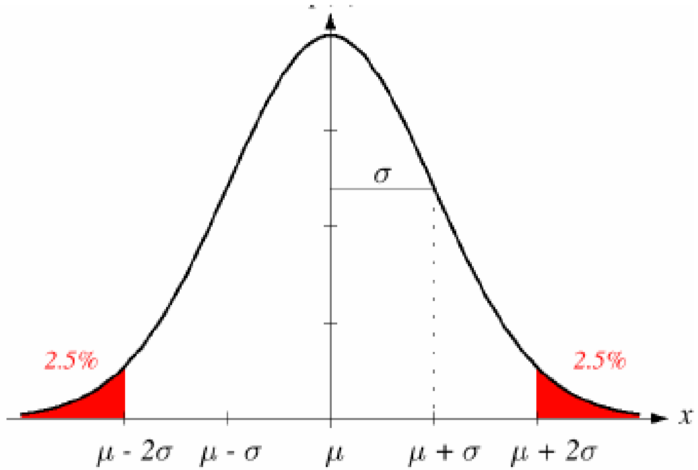
$$\left\{P(\omega_1)f(x|\omega_1)-P(\omega_2)f(x|\omega_2)\right\} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

Suppose fish lightness had a Gaussian Distribution

- Likelihood $f(x \mid \omega_i)$ are now assumed to be Gaussians with mean μ_i and variance σ^2_i

$$f(x \mid \omega_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-1}{2} \left(\frac{x-\mu_i}{\sigma_i} \right)^2 \right)$$

$f(x \mid \omega_i)$



Log Likelihood ratio

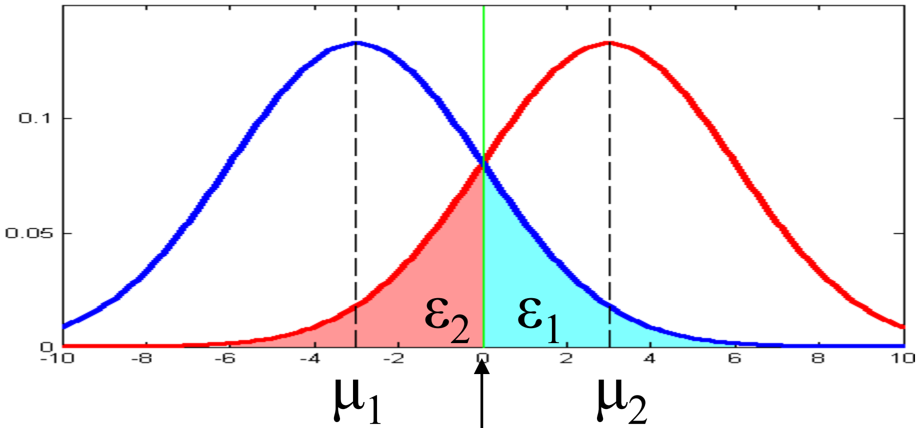
$$\ln(l(x)) = \ln \left[\frac{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2} \left(\frac{x-\mu_1}{\sigma_1} \right)^2 \right)}{\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{-1}{2} \left(\frac{x-\mu_2}{\sigma_2} \right)^2 \right)} \right]$$

$$= \ln\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 - \frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2$$

Much easier to find the decision boundary

Gaussian – Linear Classifier

CASE : $\sigma_1 = \sigma_2$



$0.5(\mu_1 + \mu_2)$ if $P(\omega_2)=P(\omega_1)$

$$\ln\left(\frac{\sigma_2}{\sigma_1}\right)+\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2$$

$$=\frac{1}{2}\left[\left(\frac{x-\mu_2}{\sigma}\right)^2-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2\right]$$

$$=\frac{1}{\sigma^2}\left[x(\mu_1-\mu_2)+\left(\frac{\mu_2^{\,2}-\mu_1^{\,2}}{2}\right)\right]$$

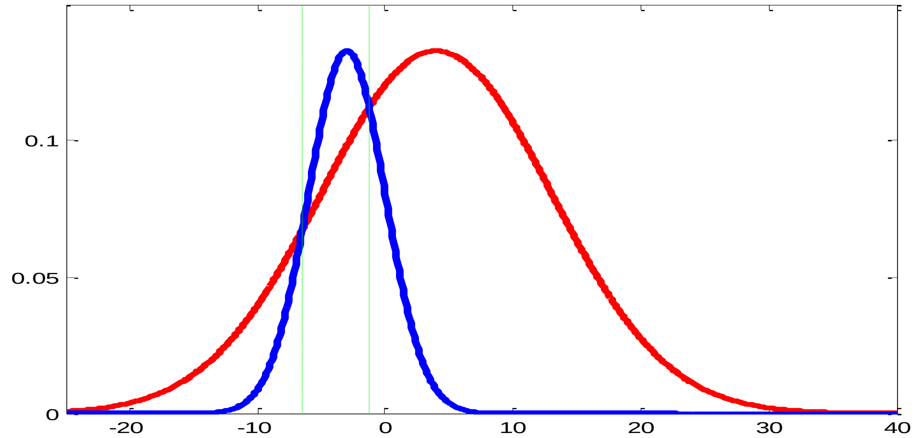
$$x(\mu_1-\mu_2)+\frac{1}{2}\left(\mu_2^{\,2}-\mu_1^{\,2}\right) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \sigma^2 \ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

If equal priors, then the threshold value for $x = 0.5(\mu_1 + \mu_2)$

$$P_e = 0.5(\varepsilon_1 + \varepsilon_2) = \varepsilon_1 = \varepsilon_2$$

Gaussian – Quadratic Classifier

CASE : $\sigma_1 \neq \sigma_2$



$$\ln\left(\frac{\sigma_2}{\sigma_1}\right)+\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

$$ax^2+bx+c \gtrless 0$$

$$(x-x_1)(x-x_2) \gtrless 0$$

Other Decision Strategies

P_e is not the only criterion we can minimize.

- Minimum Cost/Loss Classification
  - When not all errors are equally bad, and not all correct classifications are equally good. The objective criterion now is to find the smallest average cost.
- Neyman-Pearson Criterion
  - Fixes one class error probability and seeks to minimize the other.
- Minimax Criterion
  - Minimizes the maximum Bayes Risk without knowing the priors but needs a cost function.

Multiple Features

- So far, we've considered, a single feature i.e. a Gaussian distributed lightness feature for the fish

- We now consider multiple Gaussian distributed features

  e.g. lightness and weight

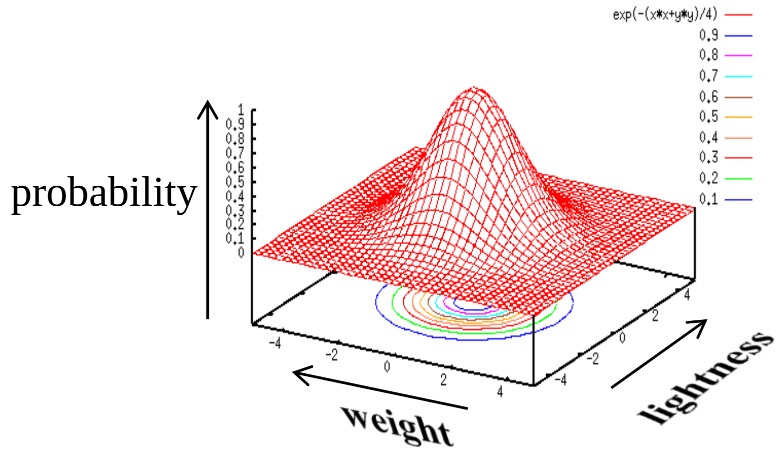
Multivariate Gaussian

$$f(\mathbf{x}\,|\,\omega_i) = \frac{1}{(2\pi)^{d/2}\,|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t\,\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right]$$

Number of dimensions

covariance matrix

For d = 2



Feature vector is now multidimensional and takes into account several features

Covariance Matrix

$$\boldsymbol{\Sigma} = E\left[(\mathbf{x}\text{-}E(\mathbf{x}))(\mathbf{x}\text{-}E(\mathbf{x}))^{\mathrm{T}}\right]$$

Symmetric, positive semi-definite

$$= \begin{bmatrix} \mathrm{cov}(x_1, x_1) & \mathrm{cov}(x_1, x_2) & \dots & \mathrm{cov}(x_1, x_d) \\ \mathrm{cov}(x_2, x_1) & \mathrm{cov}(x_2, x_2) & \dots & \mathrm{cov}(x_2, x_d) \\ \dots & & & \\ \mathrm{cov}(x_d, x_1) & \mathrm{cov}(x_d, x_2) & \dots & \mathrm{cov}(x_d, x_d) \end{bmatrix}$$

$$\mathrm{cov}(x_1, x_1) = \mathrm{var}(x_1) = \sigma_1^2$$

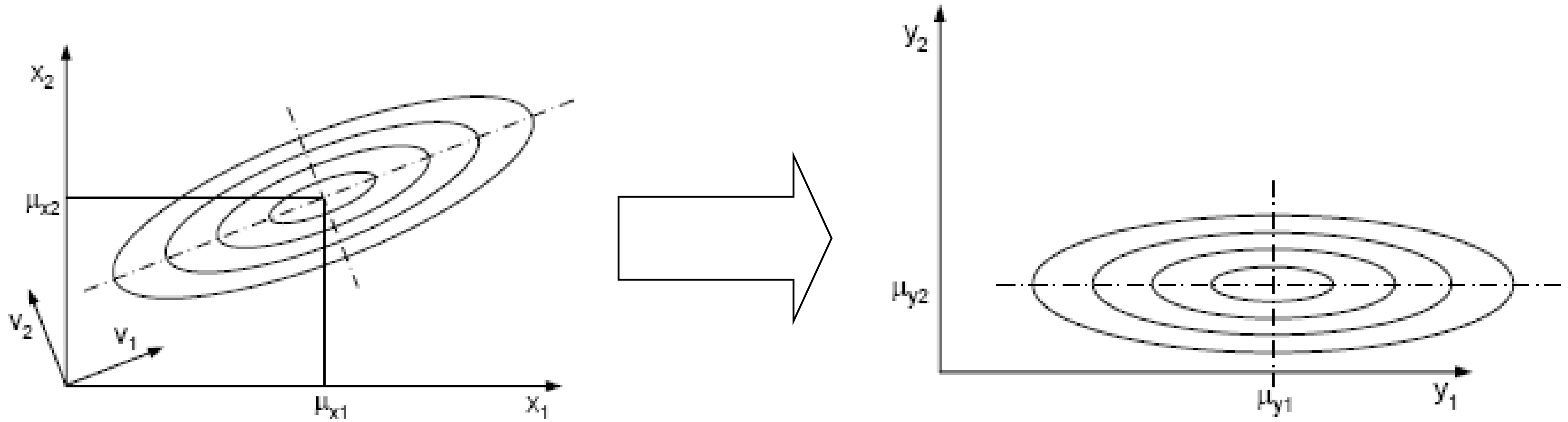
From the covariance matrix, we gain an insight on how each feature is related statistically to other features

Change of Basis - Eigenvector

Change of Basis

Given the covariance matrix Σ of a Gaussian distribution

- The **eigenvectors** of Σ are the principal directions of the distribution
- The **eigenvalues** are the variances of the corresponding principal directions

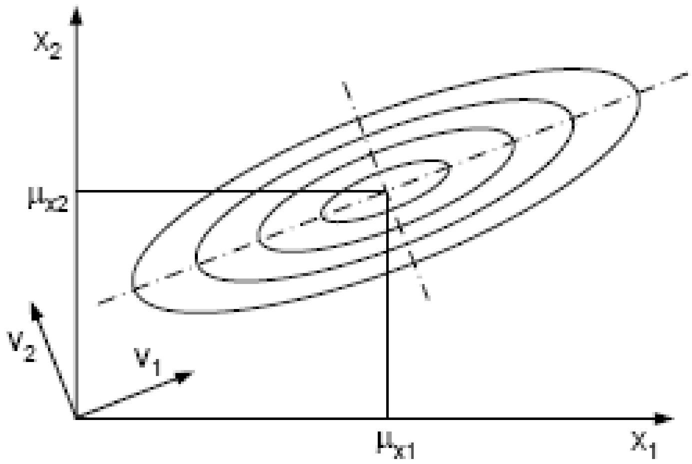


The linear transformation defined by the eigenvectors of Σ leads to basis vectors that are uncorrelated **regardless** of the form of the distribution

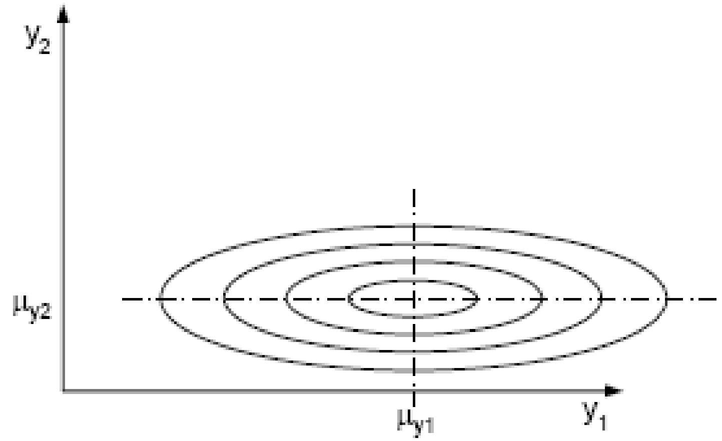
If the distribution is Gaussian, then the transformed vectors will also be statistically independent

Linear Transformation Using Eigenvectors

$$\mathbf{\Sigma V} = \mathbf{V\Lambda} \qquad \mathbf{V} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix}$$



$$\mathbf{y} = \mathbf{V}^T\mathbf{x}$$



$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\left|\mathbf{\Sigma}\right|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(y_i-\mu_{y_i})^2}{2\sigma_i^2}\right]$$

Product of univariate Gaussians

Another Detour: Diagonalization

$$\Sigma \mathbf{V} = \mathbf{V}\boldsymbol{\Lambda}$$

$$\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I} \qquad \boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{\mathbf{1/2}}\boldsymbol{\Lambda}^{\mathbf{1/2}}$$

$$\mathbf{V}^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{V} = \boldsymbol{\Lambda}$$

$$\boldsymbol{\Lambda}^{\mathbf{-1/2}}\mathbf{V}^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{V}\boldsymbol{\Lambda}^{\mathbf{-1/2}} = \mathbf{I}$$

Any square (positive definite) matrix can be diagonalized
by using its Eigenvector and Eigenvalue matrices

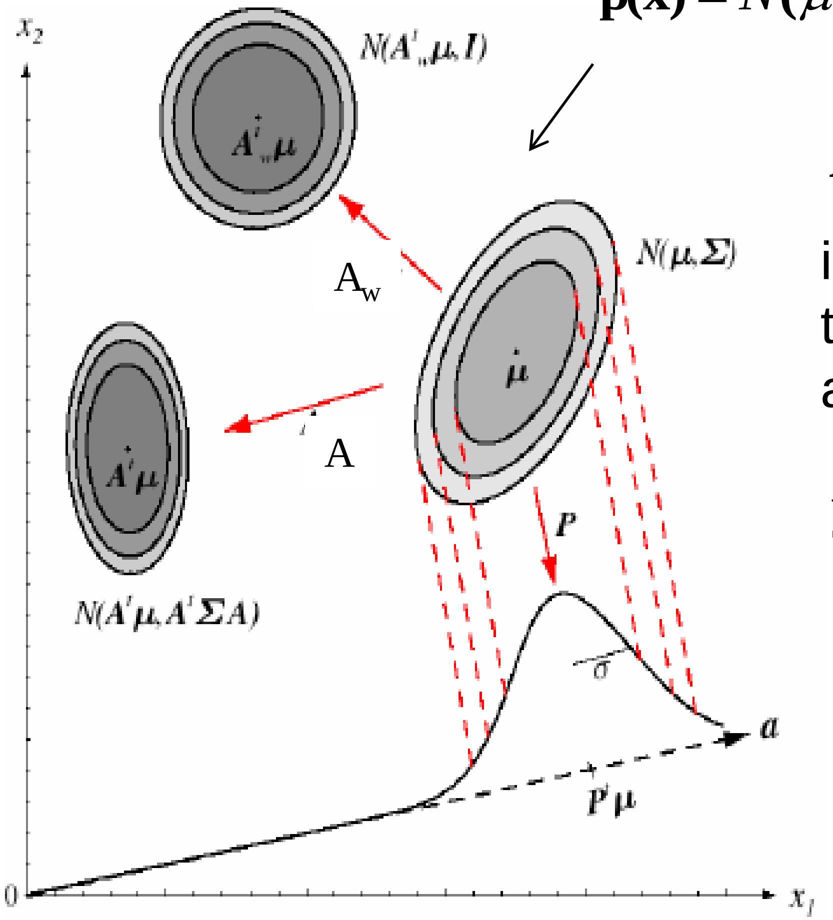
NOTE: $\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I}$ is true only if $\sum$ is symmetric and positive definite.

Otherwise in general for a square matrix, the diagonal matrix is obtained as

$$\mathrm{V}^{-1}\Sigma\mathrm{V} = \Lambda$$

Whitening Transform

$$\mathbf{p(x)} = N(\mu, \Sigma)$$



$$\mathbf{A_w} = \mathbf{V\Lambda^{-1/2}}$$

is called whitening transform, where **V** is the matrix formed by the eigenvectors of **Σ** and **Λ** is the matrix formed by eigenvalues

$$\mathbf{y} = \mathbf{A_w^T x} = N(\mathbf{A^T}\mu, \mathbf{\Lambda^{-1/2} A^T \Sigma A \Lambda^{-1/2}})$$
$$= N(\mathbf{A^T \mu, I})$$

Why? Because of this property for

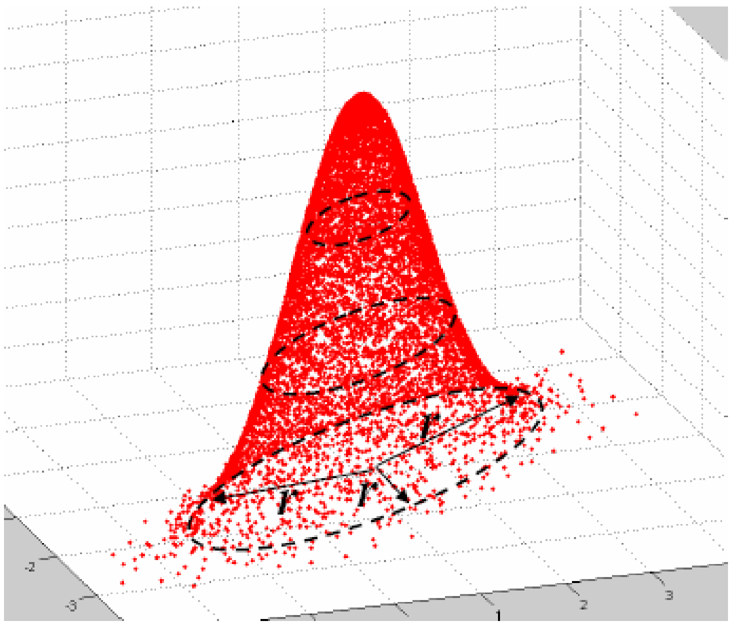
a general matrix **A**

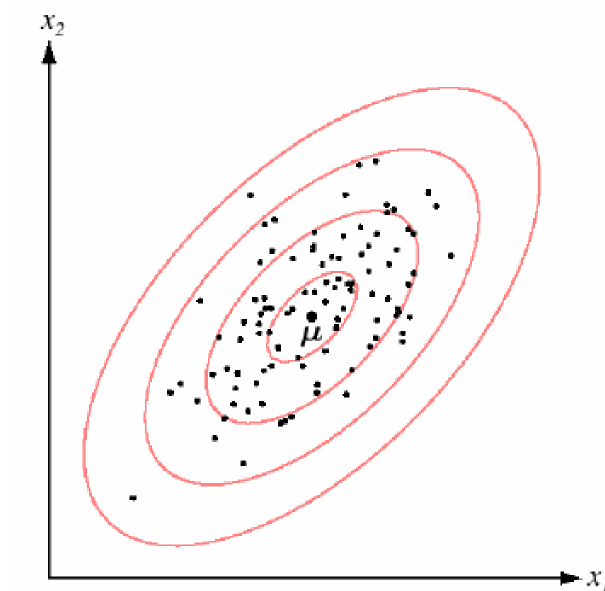
$$\mathbf{y} = \mathbf{A^T x} = N(\mathbf{A^T}\mu, \mathbf{A^T \Sigma A})$$

Mahalanobis Distance

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\,|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\underbrace{(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}\right]$$

$$r = (\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$ Square of Mahalanobis distance





Prasanta Chandra Mahalanobis (1893-1972)

r is the distance of **x** to the mean of the points normally distributed. Contours of constant density are therefore ellipsoids of constant Mahalanobis distance from the mean.

Distance Measures

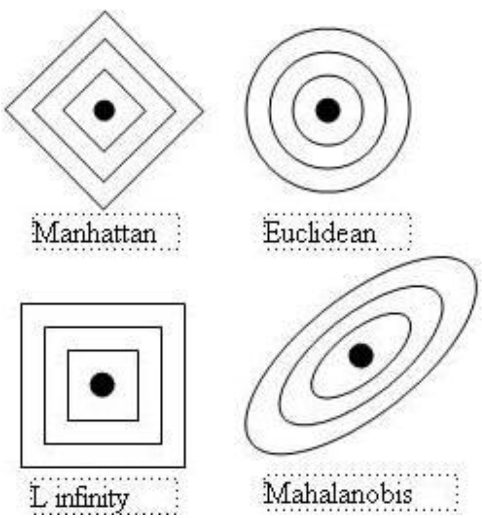
Manhattan Distance = $\|\mathbf{x}-\mathbf{y}\|_1 = \sum_{i=1}^{n} |x_i - y_i|$

Euclidian Distance = $\|\mathbf{x}-\mathbf{y}\|_2 = \sqrt{(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})}$

Mahalanobis Distance = $\|\mathbf{x}-\mathbf{y}\|_{nobis} = \sqrt{(\mathbf{x}-\mathbf{y})^T \Sigma^{-1} (\mathbf{x}-\mathbf{y})}$

Infinity Distance = $\|\mathbf{x}-\mathbf{y}\|_\infty = \lim_{p\to\infty}\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}$

$$= \max\left(|x_1 - y_1|, \ldots, |x_n - y_n|\right)$$



We should choose a proper distance measure depending on the problem