# Performance Evaluation for Classification

# Confusion Matrix – 2-class

| | | **Actual Class** | |
|---|---|---|---|
| **Predicted Class** | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive (C11) | False Positive (C12) |
| | Class2 (Negative) | False Negative (C21) | True Negative (C22) |

- True Positive: Number of test samples correctly predicted as positive class (Class1).
- True Negative: Number of test samples correctly predicted as negative class (Class2).
- False Positive: Number of test samples predicted as positive class (Class1) but actually belonging to negative class (Class2).
- False Negative: Number of test samples predicted as negative class (Class2) but actually belonging to positive class (Class1).

# Confusion Matrix – 2-class

| | | Actual Class | |
|---|---|---|---|
| Predicted Class | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive | False Positive |
| | Class2 (Negative) | False Negative | True Negative |

**Total test samples in class1**

# Confusion Matrix – 2-class

| | | **Actual Class** | |
|---|---|---|---|
| **Predicted Class** | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive | False Positive |
| | Class2 (Negative) | False Negative | True Negative |

**Total test samples in class2**

# Confusion Matrix – 2-class

| | | Actual Class | |
|---|---|---|---|
| **Predicted Class** | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive | False Positive |
| | Class2 (Negative) | False Negative | True Negative |

**Total test samples predicted as class1**

- Biometric authentication system to access account
  - False Positive (wrongly detecting as genuine person) should be low
  - Some False Negative (Not detecting a genuine person as genuine) is OK
  - Precision should be high

# Confusion Matrix – 2-class

| | | **Actual Class** | |
|---|---|---|---|
| **Predicted Class** | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive | False Positive |
| | Class2 (Negative) | False Negative | True Negative |

**Total test samples predicted as class2**

- Medical image analysis of microscopic image to detect the presence of cancer
  - False Negative (Detecting cancerous image as not cancer) should be low
  - Some False Positive (Detecting a non-cancerous images as cancer) is OK
  - Recall should be high

# Accuracy – 2-class

$$\text{Accuracy}(\%) = \frac{\text{Number of samples correctly classified } (C11 + C22)}{\text{Total number of samples used for testing}} * 100$$

$$\text{Accuracy}(\%) = \frac{TP + TN}{\text{Total number of samples used for testing}} * 100$$

| | | | |
|---|---|---|---|
| | **Actual Class** | | |
| **Predicted Class** | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive (C11) | False Positive (C12) |
| | Class2 (Negative) | False Negative (C21) | True Negative (C22) |

# Confusion Matrix - Multiclass

**Illustration:** Number of classes is 3. It can be extended to any number of classes

| | Actual Class | | | |
|---|---|---|---|---|
| **Predicted Class** | | Class1 | Class2 | Class3 |
| | Class1 | C11 | C12 | C13 |
| | Class2 | C21 | C22 | C23 |
| | Class3 | C31 | C32 | C33 |

- C11: Number of test examples predicted as class1 and actually belonging to class1
- C12: Number of test examples predicted as class1, but actually belonging to class2
- C13: Number of test examples predicted as class1, but actually belonging to class3
- *Similarly C21, C22, C23, C31, C32 and C33 are interpreted*

# Confusion Matrix - Multiclass

**With reference to Class1:**

| Actual Class | | | | |
|---|---|---|---|---|
| Predicted Class | | Class1 | Class2 | Class3 |
| | Class1 | C11 | C12 | C13 |
| | Class2 | C21 | C22 | C23 |
| | Class3 | C31 | C32 | C33 |

- True Positive: Number of test samples correctly predicted as positive class (class1) (C11).
- True Negative: Number of test samples correctly predicted as negative class (class2 and class3) (C22+C33).
- False Positive: Number of test samples predicted as positive class (class1) but actually belonging to negative class (class2 and class3) (C12+C13)
- False Negative: Number of test samples predicted as negative class (class2 and class3) but actually belonging to positive class (class1) (C21+C31)

# Confusion Matrix - Multiclass

**With reference to Class2:**

| | | Actual Class | | |
|---|---|---|---|---|
| **Predicted Class** | | Class1 | Class2 | Class3 |
| | Class1 | C11 | C12 | C13 |
| | Class2 | C21 | C22 | C23 |
| | Class3 | C31 | C32 | C33 |

- True Positive: Number of test samples correctly predicted as positive class (class2) (C22).
- True Negative: Number of test samples correctly predicted as negative class (class1 and class3) (C11+C33).
- False Positive: Number of test samples predicted as positive class (class2) but actually belonging to negative class (class1 and class3) (C21+C23)
- False Negative: Number of test samples predicted as negative class (class1 and class3) but actually belonging to positive class (class2) (C12+C32)

# Confusion Matrix - Multiclass

**With reference to Class3:**

| Actual Class | | | | |
|---|---|---|---|---|
| **Predicted Class** | | Class1 | Class2 | Class3 |
| | Class1 | C11 | C12 | C13 |
| | Class2 | C21 | C22 | C23 |
| | Class3 | C31 | C32 | C33 |

- True Positive: Number of test samples correctly predicted as positive class (class3) (C33).
- True Negative: Number of test samples correctly predicted as negative class (class1 and class2) (C11+C22).
- False Positive: Number of test samples predicted as positive class (class3) but actually belonging to negative class (class1 and class2) (C31+C32)
- False Negative: Number of test samples predicted as negative class (class1 and class2) but actually belonging to positive class (class3) (C13+C23)

# Confusion Matrix - Multiclass

**Example: Number of classes = 3. Same concept can be extended to number of classes more than 3**

| | | Actual Class | | | |
|---|---|---|---|---|---|
| **Predicted Class** | | Class1 | Class2 | Class3 | |
| | Class1 | C11 | C21 | C31 | Total samples predicted as class1 |
| | Class2 | C12 | C22 | C32 | Total samples predicted as class2 |
| | Class2 | C13 | C23 | C33 | Total samples predicted as class3 |
| **Total** | | Total samples in class1 | Total samples in class2 | Total samples in class3 | |

Total samples used for testing

# Accuracy of Multiclass Classification

**Example: Number of classes = 3. Same concept can be extended to number of classes more than 3**

$$\text{Accuracy}(\%) = \frac{\text{Number of samples correctly classified } (C11 + C22 + C33)}{\text{Total number of samples used for testing}} * 100$$

$$\text{Accuracy}(\%) = \frac{TP + TN}{\text{Total number of samples used for testing}} * 100$$

| | | Actual Class | | |
|---|---|---|---|---|
| **Predicted Class** | | Class1 | Class2 | Class3 |
| | Class1 | C11 | C21 | C31 |
| | Class2 | C12 | C22 | C32 |
| | Class2 | C13 | C23 | C33 |

# Binary (2-class) Classification: Precision, Recall and F-measure

| | | Actual Class | |
|---|---|---|---|
| | | Class1 (Positive) | Class2 (Negative) |
| Predicted Class | Class1 (Positive) | True Positive (TP) | False Positive (FP) |
| | Class2 (Negative) | False Negative (FN) | True Negative (TN) |

- Precision:
  - Number of samples correctly classified as positive class, out of all the examples classified as positive class
  - It is also called positive predictive value

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Precision} = \frac{\text{Number of samples correctly classified as positive class}}{\text{Total number of samples classified as positive class}}$$

# Binary (2-class) Classification: Precision, Recall and F-measure

| | | Actual Class | |
|---|---|---|---|
| Predicted Class | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive (TP) | False Positive (FP) |
| | Class2 (Negative) | False Negative (FN) | True Negative (TN) |

- Recall:
  - Number of samples correctly classified as positive class, out of all the examples belonging to positive class
  - It is also called as sensitivity or true positive rate (TPR)

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{Number of samples correctly classified as positive class}}{\text{Total number of samples belonging to positive class}}$$

# Binary (2-class) Classification: Precision, Recall and F-measure

| | | Actual Class | |
|---|---|---|---|
| Predicted Class | | Class1 (Positive) | Class2 (Negative) |
| | Class1 (Positive) | True Positive (TP) | False Positive (FP) |
| | Class2 (Negative) | False Negative (FN) | True Negative (TN) |

- F-measure or F-score or F1-score:
  - Combines precision and recall
  - Recall and precision are evenly weighted.
  - Harmonic mean of precision and recall

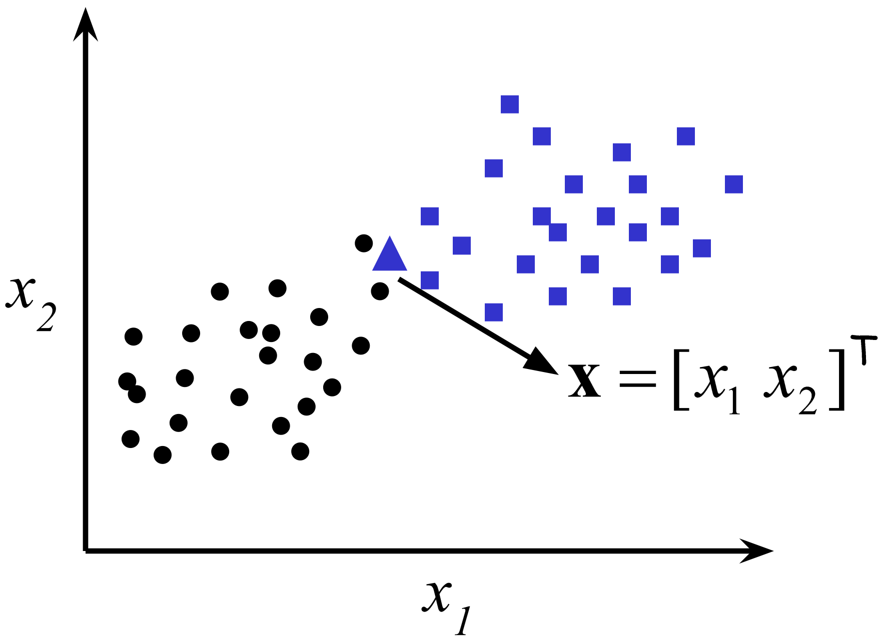
$$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

# Supervised Machine Learning: Pattern Classification

## K-Nearest Neighbor, Reference Template Method

# K-Nearest Neighbours (K-NN) Method

- Consider the class labels of the K training examples nearest to the test example
- Step 1: Compute Euclidean distance for a test example $\mathbf{x}$ with every training examples, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_N$




- Step 2: Sort the examples in the training set in the ascending order of the distance to $\mathbf{x}$
- Step 3: Choose the first K examples in the sorted list
  - K is the number of neighbours for text example
- Step 4: Test example is assigned the most common class among its K neighbours

# Reference Templates Method

- Each class is represented by its reference templates
  - Mean of each data points of each class as reference template
  - Let the data of class $i$ be $\mathrm{D}_i = \{\mathbf{x}_n\}_{n=1}^{N_i}, \mathbf{x}_n \in \mathrm{R}^d$
    - $N_i$: Number of examples (data points) in class $i$
  - Mean of data points of a class $i$, $\boldsymbol{\mu}_i$ is given as:

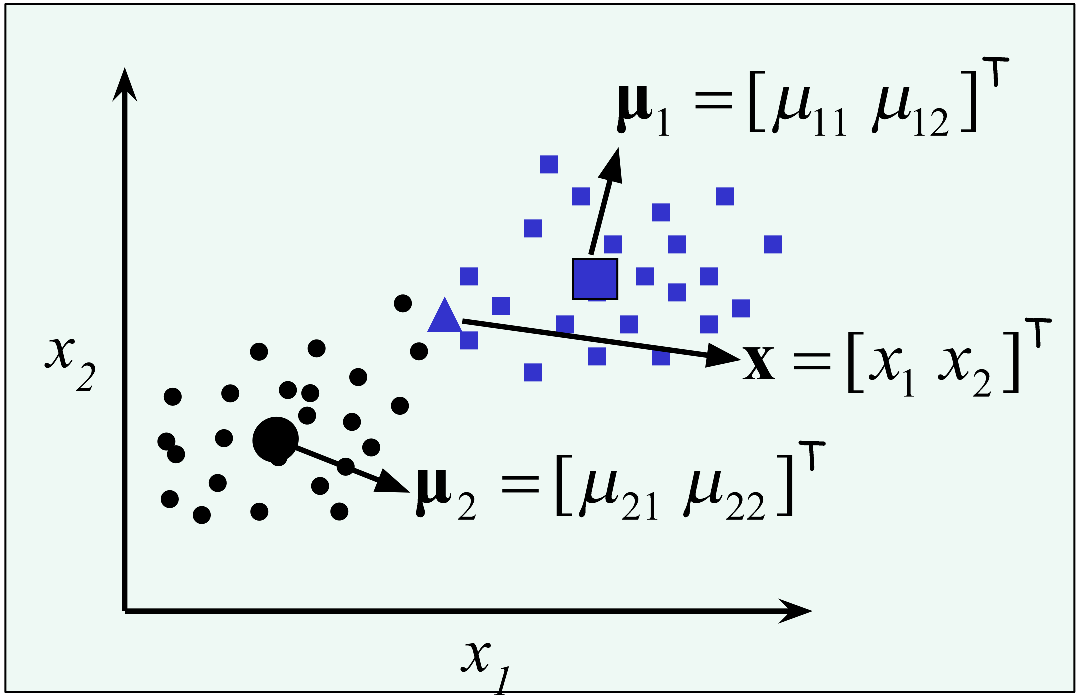
$$\boldsymbol{\mu}_i = \frac{1}{N_i}\sum_{n=1}^{N_i}\mathbf{x}_n$$

# Reference Templates Method

- Each class is represented by its reference templates
  - Mean of each data points of each class as reference template
- For a test example, compute an Euclidean distance to all the reference template corresponding to each class, $ED(\mathbf{x}, \boldsymbol{\mu}_i)$

$$= \text{argmin } ED(\mathbf{x}, \boldsymbol{\mu}_i)$$

$\boldsymbol{\mu}_i$: Mean vector of class $i$




- The class of the nearest reference template (mean) is assigned to the test pattern

Class label for $\mathbf{x} = \underset{i}{\text{argmin}}\ ED(\mathbf{x}, \boldsymbol{\mu}_i)$

$$i = 1,2,\ldots, M$$

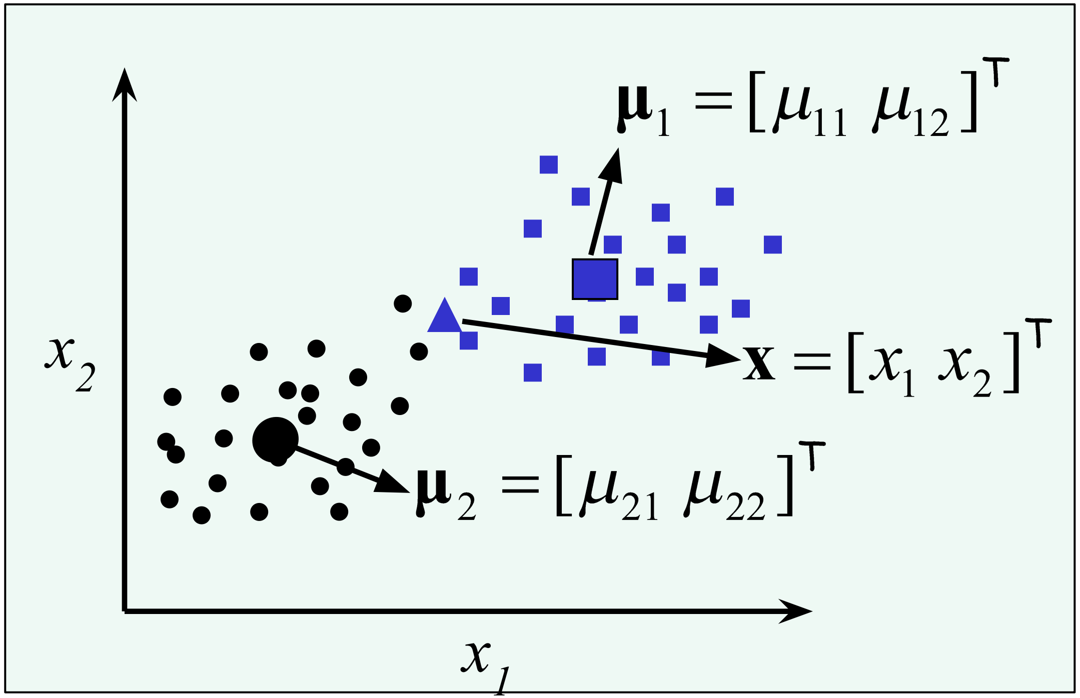
$M$ = Number of classes

# Reference Templates Method

- Each class is represented by its reference templates
  - Mean of each data points of each class as reference template
- For a test example, compute an Euclidean distance to all the reference template corresponding to each class, $ED(\mathbf{x}, \boldsymbol{\mu}_i)$

$$= \operatorname{argmin} ED(\mathbf{x}, \boldsymbol{\mu}_i)$$

$\boldsymbol{\mu}_i$: Mean vector of class $i$




- The class of the nearest reference template (mean) is assigned to the test pattern
- Learning: Estimating first order statistics (mean) from the data of each class

# Illustration of Reference Templates Method: Adult(1)-Child(0) Classification

| Height | Weight | Class |
|---|---|---|
| 90 | 21.5 | 0 |
| 95 | 23.67 | 0 |
| 100 | 32.45 | 0 |
| 116 | 38.21 | 0 |
| 98 | 28.43 | 0 |
| 108 | 36.32 | 0 |
| 104 | 27.38 | 0 |
| 112 | 39.28 | 0 |
| 121 | 35.8 | 0 |
| 92 | 23.56 | 0 |
| 152 | 46.8 | 1 |
| 178 | 78.9 | 1 |
| 163 | 67.45 | 1 |
| 173 | 82.9 | 1 |
| 154 | 52.6 | 1 |
| 168 | 66.2 | 1 |
| 183 | 90 | 1 |
| 172 | 82 | 1 |
| 156 | 45.3 | 1 |
| 161 | 59 | 1 |

- Training Phase:
  - Compute sample mean vector from training data of class 0 (Child)

$$\boldsymbol{\mu}_0 = [103.60 \quad 30.66]$$

# Illustration of Reference Templates Method: Adult(1)-Child(0) Classification

| Height | Weight | Class |
|---|---|---|
| 90 | 21.5 | 0 |
| 95 | 23.67 | 0 |
| 100 | 32.45 | 0 |
| 116 | 38.21 | 0 |
| 98 | 28.43 | 0 |
| 108 | 36.32 | 0 |
| 104 | 27.38 | 0 |
| 112 | 39.28 | 0 |
| 121 | 35.8 | 0 |
| 92 | 23.56 | 0 |
| 152 | 46.8 | 1 |
| 178 | 78.9 | 1 |
| 163 | 67.45 | 1 |
| 173 | 82.9 | 1 |
| 154 | 52.6 | 1 |
| 168 | 66.2 | 1 |
| 183 | 90 | 1 |
| 172 | 82 | 1 |
| 156 | 45.3 | 1 |
| 161 | 59 | 1 |

- Training Phase:
  - Compute sample mean vector from training data of class 0 (Child)

    $$\boldsymbol{\mu}_0 = [103.60 \ \ 30.66]$$

  - Compute sample mean vector from training data of class 1 (Adult)

    $$\boldsymbol{\mu}_1 = [166.00 \ \ 67.12]$$

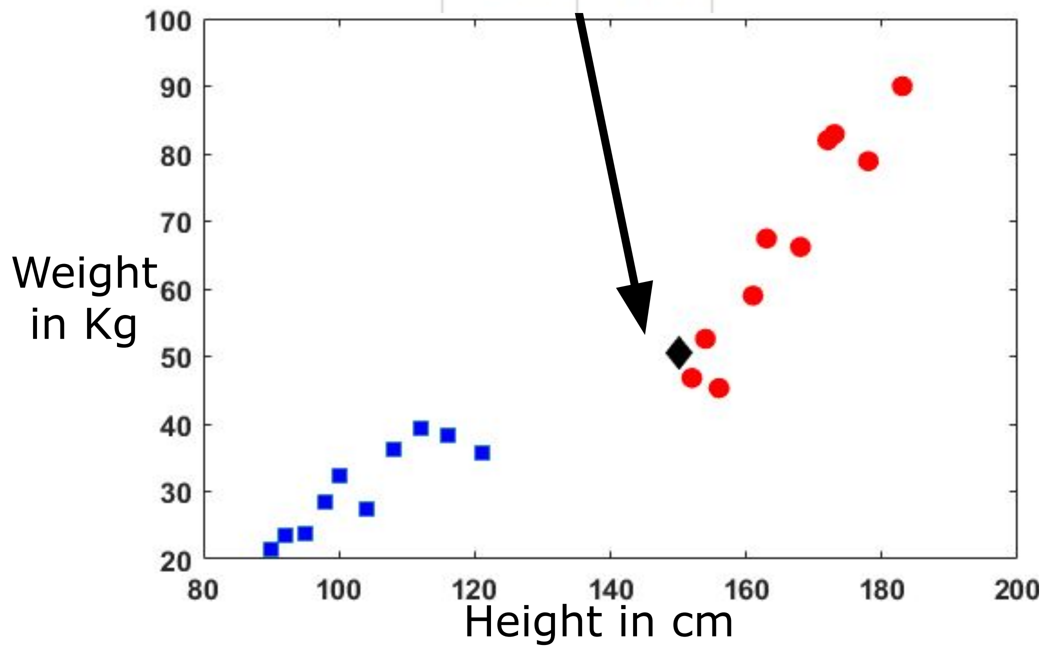
# Illustration of Reference Templates Method: Adult(1)-Child(0) Classification

- Test Phase - Classification:

**Test Example, $\mathbf{x}$ :**

| | |
|---|---|
| 150 | 50.6 |

| | Height | Weight | Class |
|---|---|---|---|
| $\boldsymbol{\mu}_0$ | 103.60 | 30.66 | 0 |
| $\boldsymbol{\mu}_1$ | 166.60 | 67.12 | 1 |




- Compute Euclidean distance of test sample, $\mathbf{x}$ with mean vector of class 0 (Child), $\boldsymbol{\mu}_0$: $ED(\mathbf{x}, \boldsymbol{\mu}_0) = 50.50$
- Compute Euclidean distance of test sample, $\mathbf{x}$ with mean vector of class 1 (Adult), $\boldsymbol{\mu}_1$: $ED(\mathbf{x}, \boldsymbol{\mu}_1) = 23.00$

Class label of $\mathbf{x}$ = Adult

# Modified Reference Templates Method

- Each class is represented by its reference templates
  - Mean and variance (covariance) of data points of each class as reference template
  - Let the data of class $i$ be $\mathrm{D}_\mathrm{i} = \{\mathbf{x}_n\}_{n=1}^{N_i}, \mathbf{x}_n \in \mathrm{R}^d$
    - $N_i$: Number of examples (data points) in class $i$
  - Mean of data points of a class $i$, $\boldsymbol{\mu}_i$ is given as:

$$\boldsymbol{\mu}_i = \frac{1}{N_i}\sum_{n=1}^{N_i}\mathbf{x}_n$$

  - Covariance matrix of data points of a class $i$, $\boldsymbol{\Sigma}_i$ is given as:

$$\Sigma_i = \frac{1}{N_i - 1}\sum_{n=1}^{N_i}(\mathbf{x}_n - \boldsymbol{\mu}_i)(\mathbf{x}_n - \boldsymbol{\mu}_i)^\mathsf{T}$$

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_1^2 \ \sigma_{12} \dots \sigma_{1d} \\ \sigma_{21}\ \sigma_2^2 \ \dots \sigma_{2d} \\ \dots\dots \\ \sigma_{d1}\ \sigma_{d2} \dots \sigma_d^2 \end{bmatrix}$$

$\sigma_j^2$ is variance; $\sigma_j^2 = \frac{1}{N_i - 1}\sum_{n=1}^{N_i}(x_{nj} - \mu_{ji})^2$

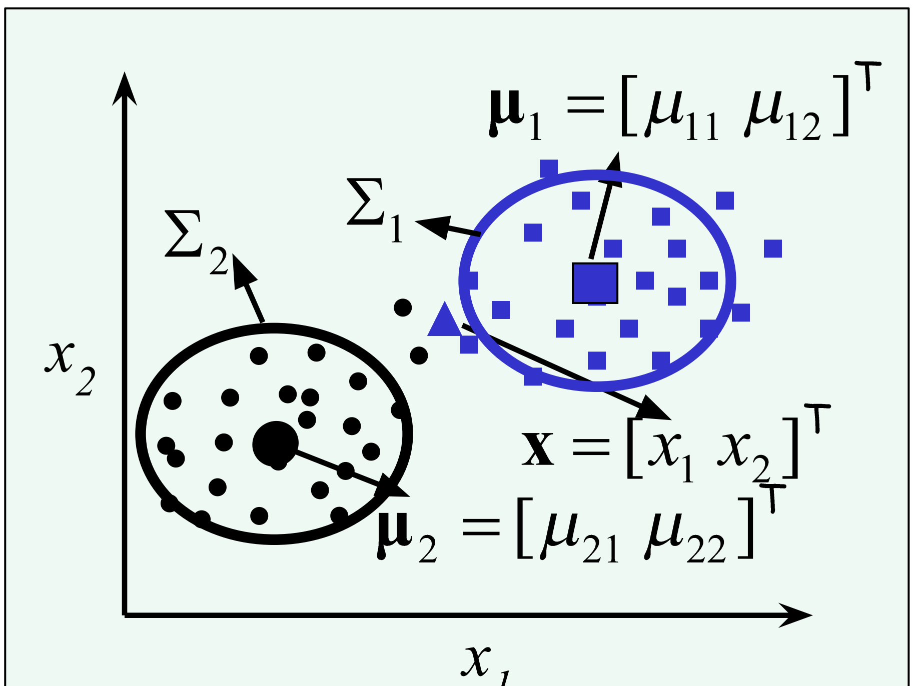
$\sigma_{jk}$: Covariance of $j^{\text{th}}$ and $k^{\text{th}}$ attribute $\sigma_{jk} = \frac{1}{N_i - 1}\sum_{n=1}^{N_i}(x_{nj} - \mu_{ij})(x_{nk} - \mu_{ik})$

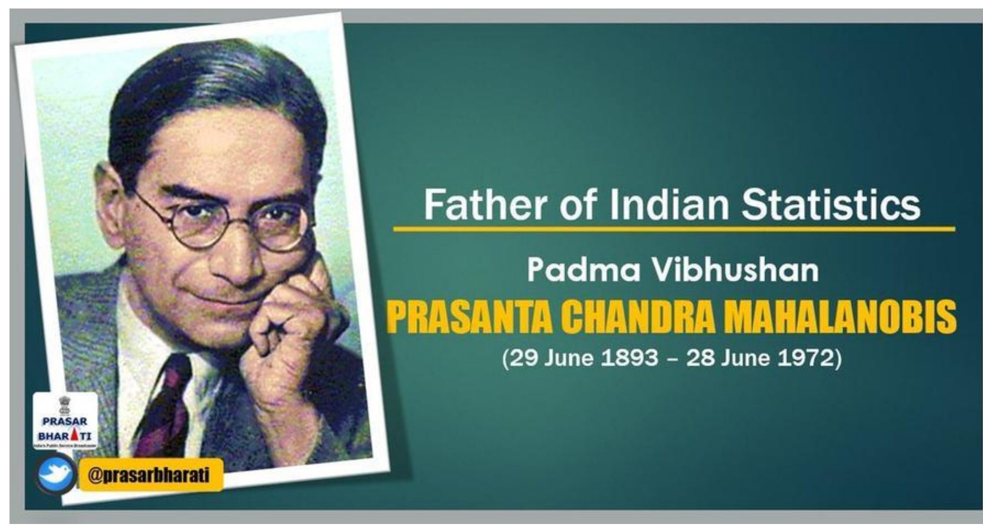
# Modified Reference Templates Method

- Each class is represented by one or more reference templates
  - Mean and variance (covariance) of data points of each class as reference template
- For a test example, compute a **Mahalanobis distance** to all the reference template corresponding to each class, $MD(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$

$$= \operatorname{argmin} MD(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

$\boldsymbol{\mu}_i$ & $\boldsymbol{\Sigma}_i$ : Mean vector and Covariance matrix of class $i$



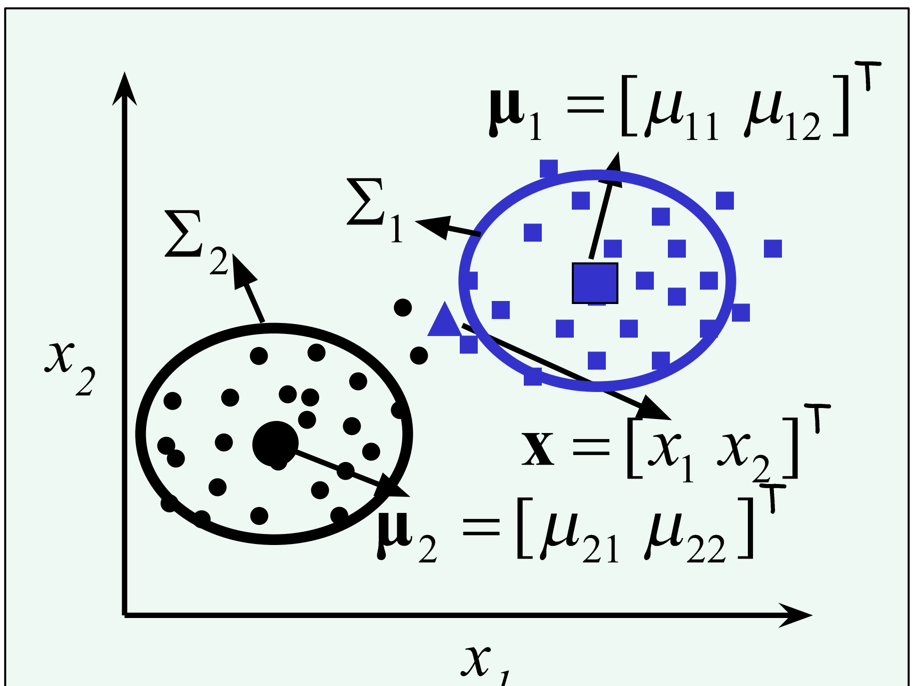

The **Mahalanobis distance** is a measure of the distance between a point and a distribution

# Modified Reference Templates Method

- Each class is represented by one or more reference templates
  - Mean and variance (covariance) of data points of each class as reference template
- For a test example, compute a **Mahalanobis distance** to all the reference template corresponding to each class, $MD(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$

$$= \operatorname{argmin} MD(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \operatorname{argmin} \sqrt{(\mathrm{x} - \mu_i)^T \Sigma_i^{-1} (\mathrm{x} - \mu_i)}$$

$\boldsymbol{\mu}_i$ & $\boldsymbol{\Sigma}_i$ : Mean vector and Covariance matrix of class $i$




- The class of the nearest reference templates is assigned to the test pattern

Class label for $\mathbf{x} = \underset{i}{\operatorname{argmin}} MD(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$

$$i = 1,2,\ldots, M$$

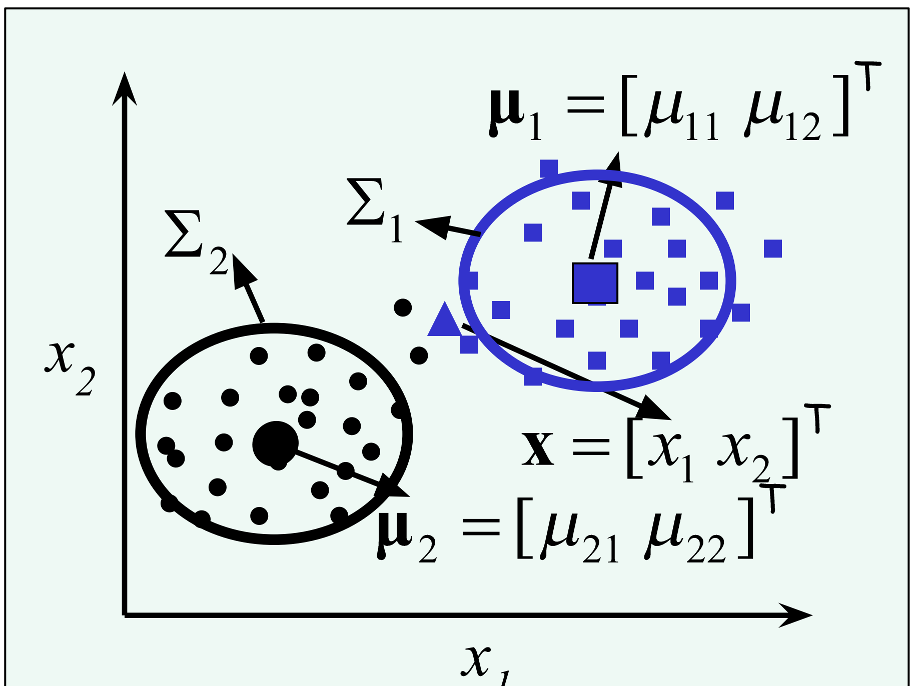
$M$ = Number of classes

# Modified Reference Templates Method

- Each class is represented by one or more reference templates
  - Mean and variance (covariance) of data points of each class as reference template
- For a test example, compute a **Mahalanobis distance** to all the reference template corresponding to each class, $MD(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$

$$= \operatorname{argmin} MD(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad = \operatorname{argmin} \sqrt{(\mathrm{x} - \mu_i)^T \Sigma_i^{-1} (\mathrm{x} - \mu_i)}$$

$\boldsymbol{\mu}_i$ & $\boldsymbol{\Sigma}_i$ : Mean vector and Covariance matrix of class $i$




- The class of the nearest reference templates is assigned to the test pattern
- Learning: Estimating
  - first order statistics (mean) and
  - Second order statistics (variance and covariance) from the data of each class

# Illustration of Reference Templates Method: Adult(1)-Child(0) Classification

| Height | Weight | Class |
|---|---|---|
| 90 | 21.5 | 0 |
| 95 | 23.67 | 0 |
| 100 | 32.45 | 0 |
| 116 | 38.21 | 0 |
| 98 | 28.43 | 0 |
| 108 | 36.32 | 0 |
| 104 | 27.38 | 0 |
| 112 | 39.28 | 0 |
| 121 | 35.8 | 0 |
| 92 | 23.56 | 0 |
| 152 | 46.8 | 1 |
| 178 | 78.9 | 1 |
| 163 | 67.45 | 1 |
| 173 | 82.9 | 1 |
| 154 | 52.6 | 1 |
| 168 | 66.2 | 1 |
| 183 | 90 | 1 |
| 172 | 82 | 1 |
| 156 | 45.3 | 1 |
| 161 | 59 | 1 |

- Training Phase:
  - Compute sample mean vector from training data of class 0 (Child)

$$\boldsymbol{\mu}_0 = [103.60 \quad 30.66]$$

  - Compute sample covariance matrix from training data of class 0 (Child)

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 109.38 & 61.35 \\ 61.35 & 43.54 \end{pmatrix}$$

# Illustration of Reference Templates Method: Adult(1)-Child(0) Classification

| Height | Weight | Class |
|---|---|---|
| 90 | 21.5 | 0 |
| 95 | 23.67 | 0 |
| 100 | 32.45 | 0 |
| 116 | 38.21 | 0 |
| 98 | 28.43 | 0 |
| 108 | 36.32 | 0 |
| 104 | 27.38 | 0 |
| 112 | 39.28 | 0 |
| 121 | 35.8 | 0 |
| 92 | 23.56 | 0 |
| 152 | 46.8 | 1 |
| 178 | 78.9 | 1 |
| 163 | 67.45 | 1 |
| 173 | 82.9 | 1 |
| 154 | 52.6 | 1 |
| 168 | 66.2 | 1 |
| 183 | 90 | 1 |
| 172 | 82 | 1 |
| 156 | 45.3 | 1 |
| 161 | 59 | 1 |

- Training Phase:
  - Compute sample mean vector from training data of class 0 (Child)

    $$\boldsymbol{\mu}_0 = [103.60 \quad 30.66]$$

  - Compute sample covariance matrix from training data of class 0 (Child)

    $$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 109.38 & 61.35 \\ 61.35 & 43.54 \end{pmatrix}$$

  - Compute sample mean vector from training data of class 1 (Adult)

    $$\boldsymbol{\mu}_1 = [166.00 \quad 67.12]$$

  - Compute sample covariance matrix from training data of class 1 (Adult)

    $$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 110.67 & 160.53 \\ 160.53 & 255.49 \end{pmatrix}$$

# Illustration of Reference Templates Method: Adult(1)-Child(0) Classification

- Test Phase - Classification:

$\boldsymbol{\mu}_0 = [103.60 \;\; 30.66]$

$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 109.38 & 61.35 \\ 61.35 & 43.54 \end{pmatrix}$

| Class |
|---|
| 0 |

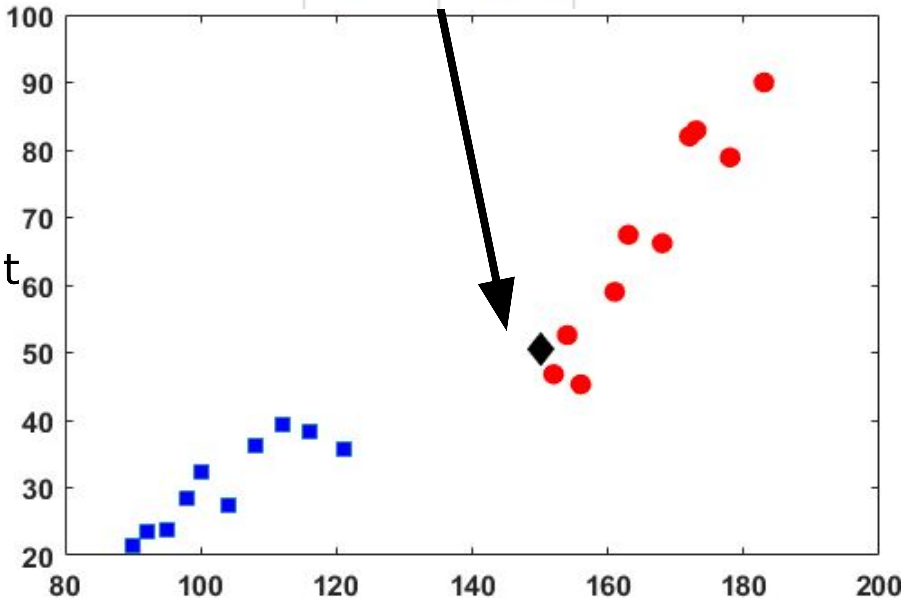
$\boldsymbol{\mu}_1 = [166.00 \;\; 67.12]$

$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 110.67 & 160.53 \\ 160.53 & 255.49 \end{pmatrix}$

| Class |
|---|
| 1 |

**Test Example, $\mathbf{x}$ :**

| | |
|---|---|
| 150 | 50.6 |




- Compute Mahalanobis distance of test sample, $\mathbf{x}$ with mean vector and covariance matrix of class 0 (Child): $MD(\mathbf{x}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) = 4.87$
- Compute Mahalanobis distance of test sample, $\mathbf{x}$ with mean vector and covariance matrix of class 1 (Adult): $MD(\mathbf{x}, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = 2.07$

Class label of $\mathbf{x}$ = Adult

# Classification using Reference Template Methods

- For a test example, a distance measure is computed with the reference template of each class
- The class of the reference template with least distance is assigned to the test pattern
- When Mahalanobis distance is used, it gives the notion that distance measure is computed between a test example and the distribution (density) of a class
  - Distribution (density) of class: All the training examples are drawn from that distribution
  - Density here is normal (Gaussian) density
- In other way, we are interested to estimate probability of class, $P(C_i \mid \mathbf{x})$
  - Given the test example $\mathbf{x}$, what is the probability that it belongs to $i^{\text{th}}$ class ($C_i$)
- Solution: Bayes classifier

# Text Books


1. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Third Edition, Morgan Kaufmann Publishers, 2011.

2. S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2009.

3. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.