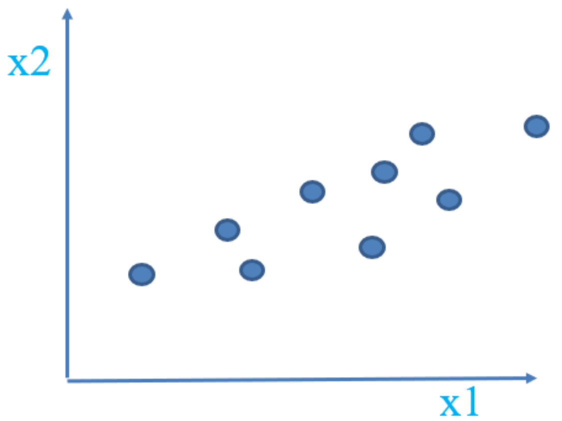
Redução de Dimensionalidade – PCA e LDA

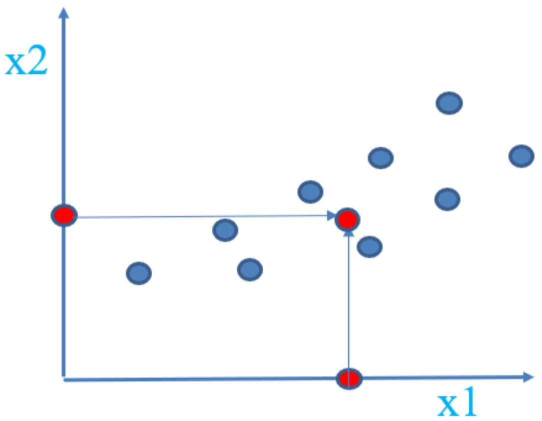
Prof. Gustavo Willam Pereira

INSTITUTO FEDERAL
Sudeste de Minas Gerais

PCA - Análise de Componentes Principais

- Considere um modelo de ML com 2 variáveis (X1, X2).
- Considere agora a média desses 2 valores de X1 e X2. E plotando o ponto médio no gráfico.

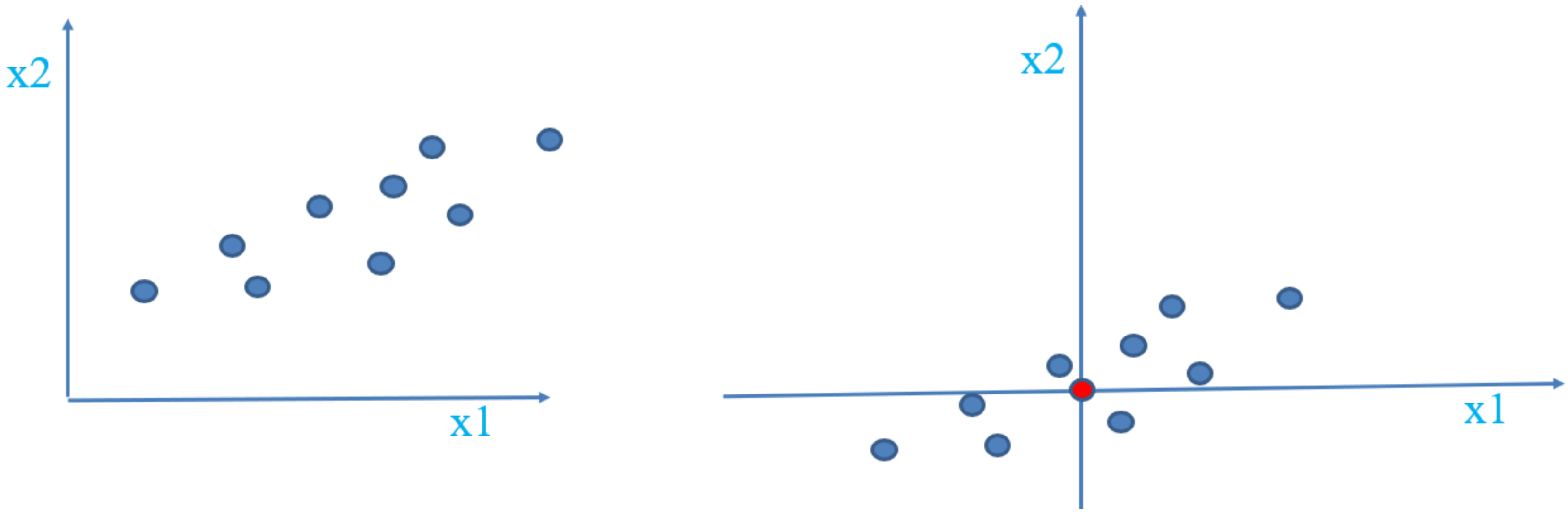




INSTITUTO FEDERAL
Sudeste de Minas Gerais

PCA - Análise de Componentes Principais

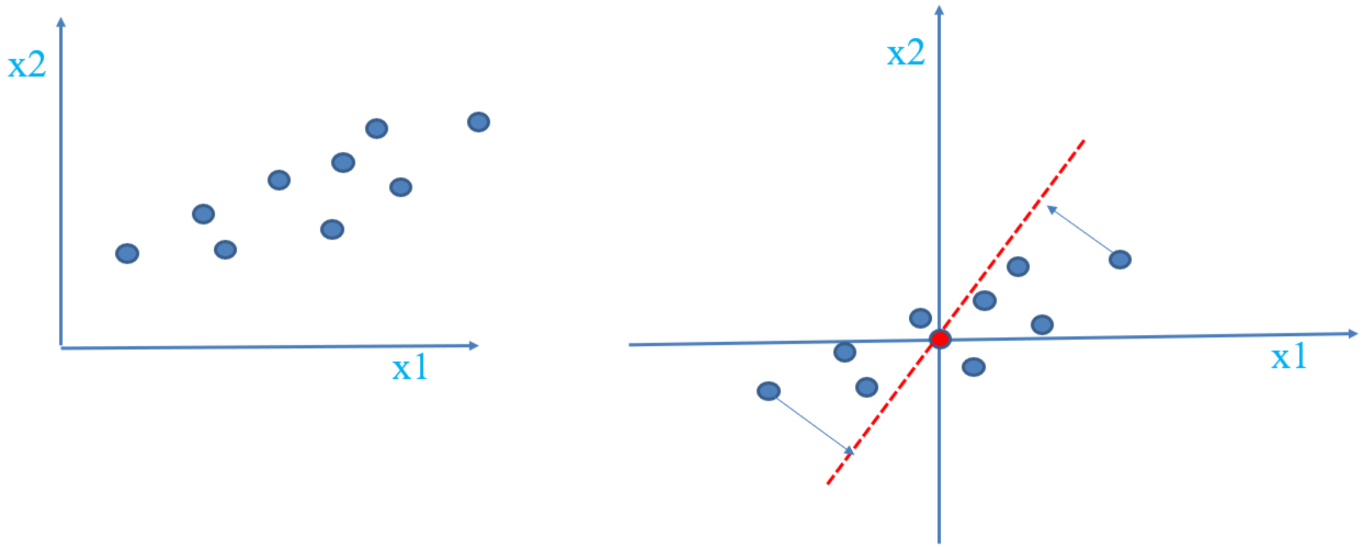
- Considere agora deslocar esses dados para a origem considerando média zero para X1 e X2



INSTITUTO FEDERAL
Sudeste de Minas Gerais

PCA - Análise de Componentes Principais

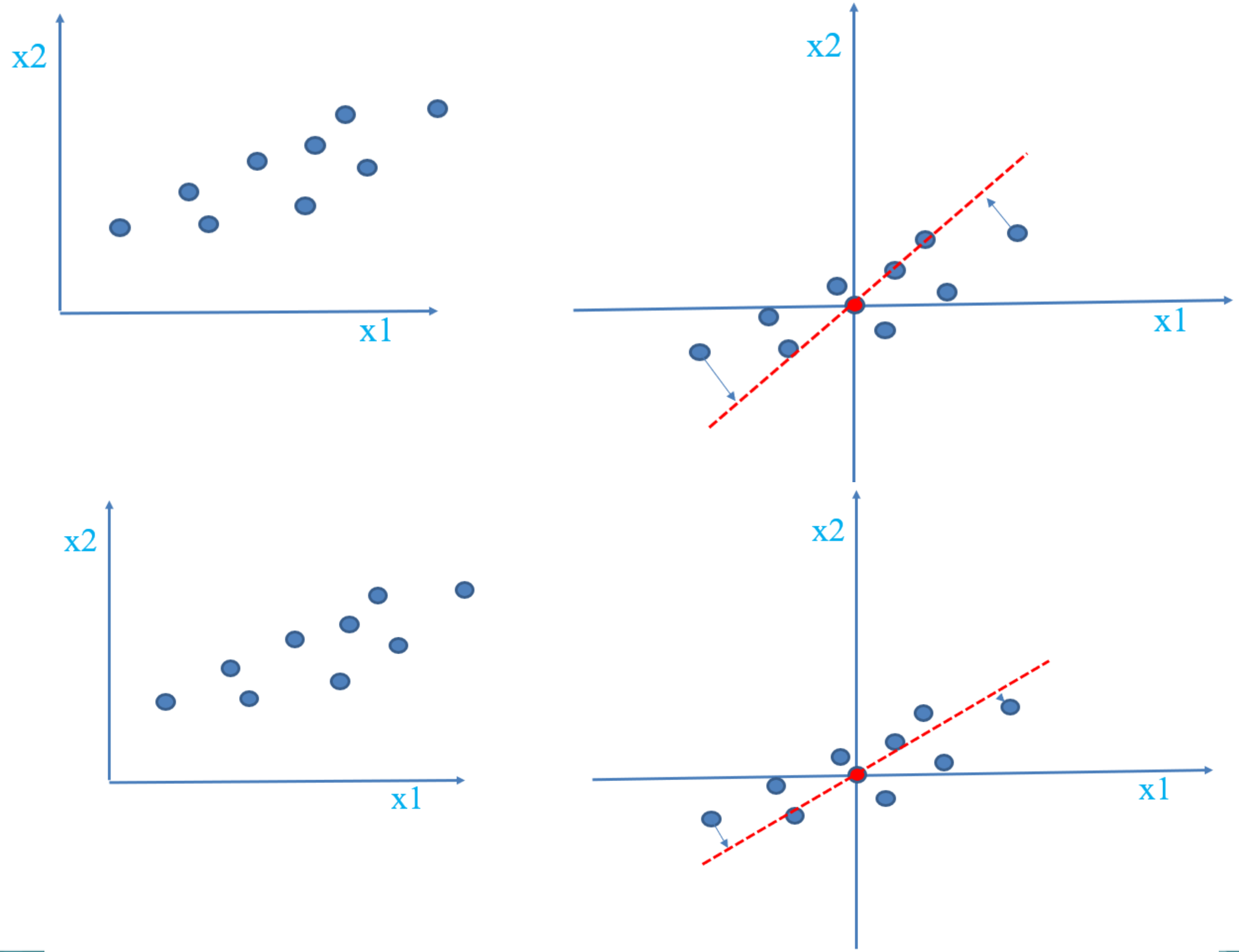
- Vamos criar um modelo linear passando entre os pontos do gráfico.



- Projetando as distâncias de todos os pontos até a reta poderíamos obter os desvios (erros).

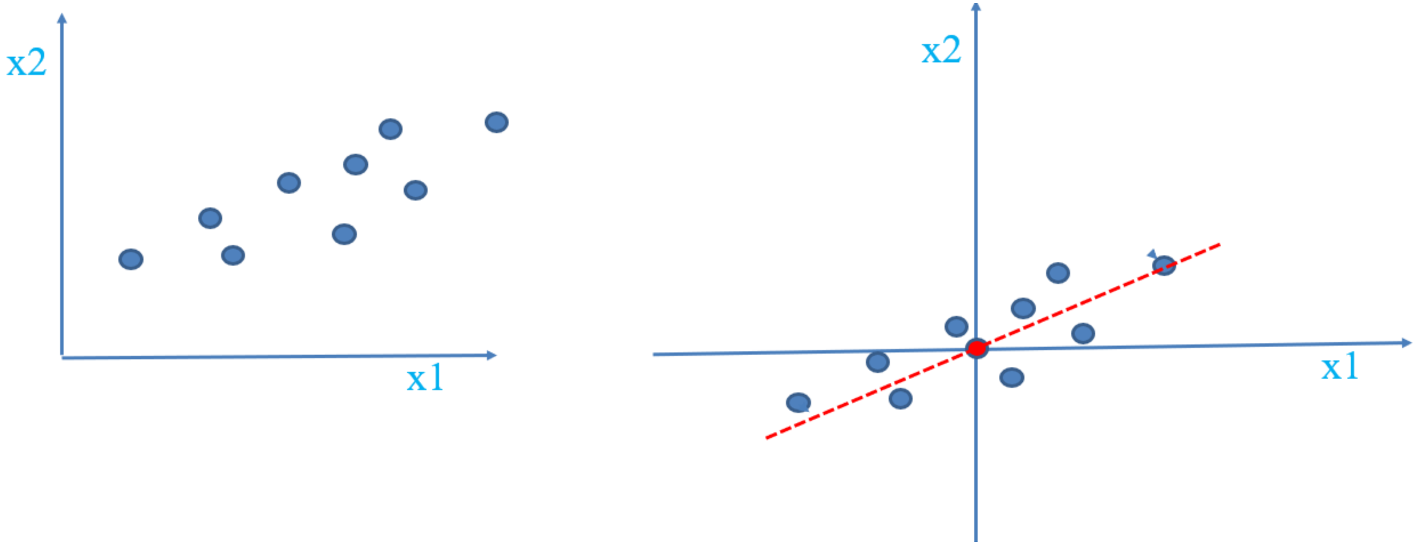
PCA - Análise de Componentes Principais

- Minimizando as distâncias entre os pontos do gráfico e a reta teríamos.



PCA - Análise de Componentes Principais

- Minimizando as distâncias entre os pontos do gráfico e a reta teríamos.



- Minimizando as distâncias até encontrar a distância mínima para os pontos.

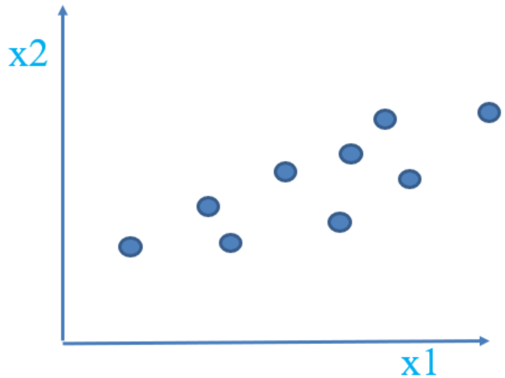
INSTITUTO FEDERAL
Sudeste de Minas Gerais

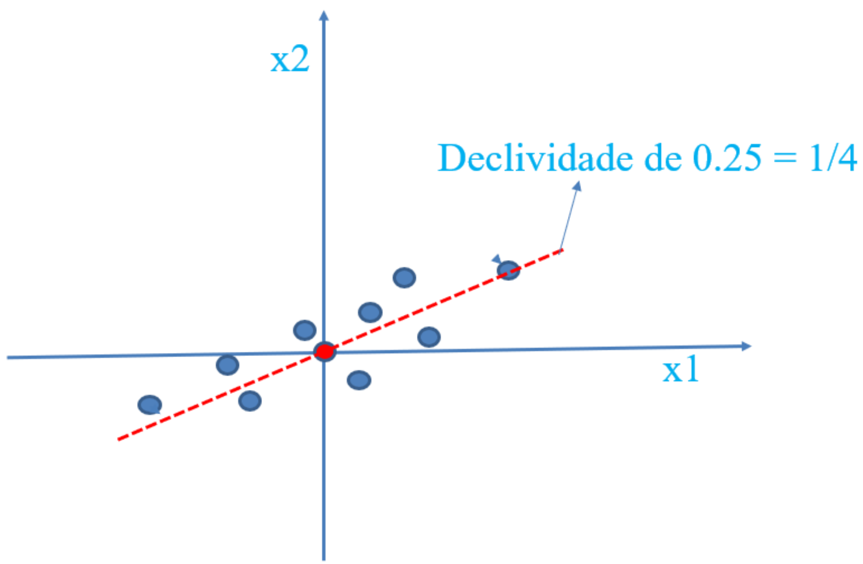
PCA - Análise de Componentes Principais

- Imagine que a declividade seja de 0,25.
- Ou seja para cada unidade deslocada na X1 eu teria que deslocar 0,25 unidades na X2.
- Poderíamos ter uma nova variável PC1, que é uma combinação linear entre X1 e X2

Pc1 = 4.x1 + 1.x2 (combinação linear)

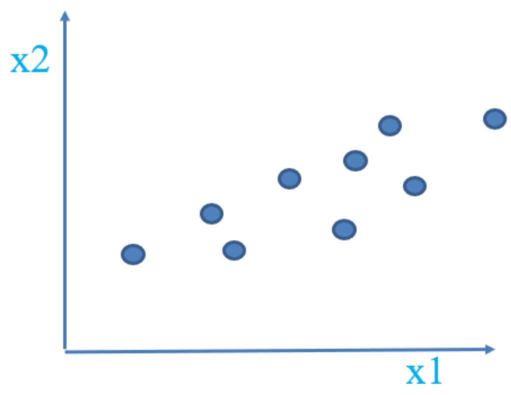
Pc1 = 0,97x1 + 0,242x2 (combinação linear)

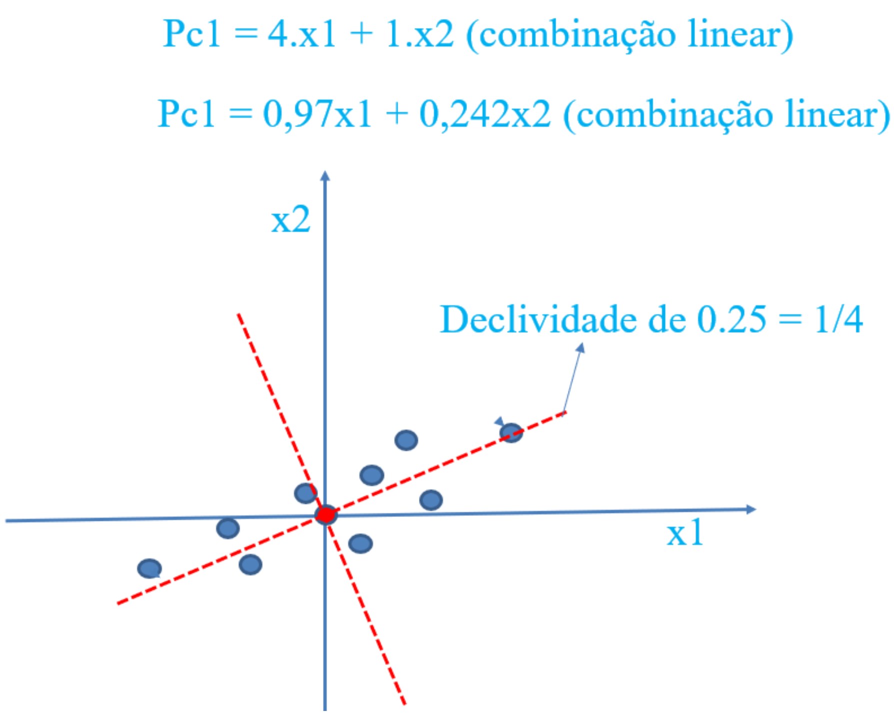




PCA - Análise de Componentes Principais

- Veja que PC1 explica uma determinada variação do modelo.
- Poderiamos ter um outro modelo linear (uma reta perpendicular a PC1), que explicaria uma certa variabilidade entre as variáveis X1 e X2.





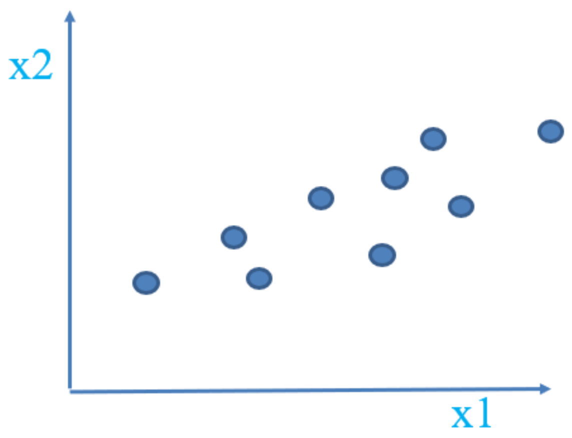
PCA - Análise de Componentes Principais

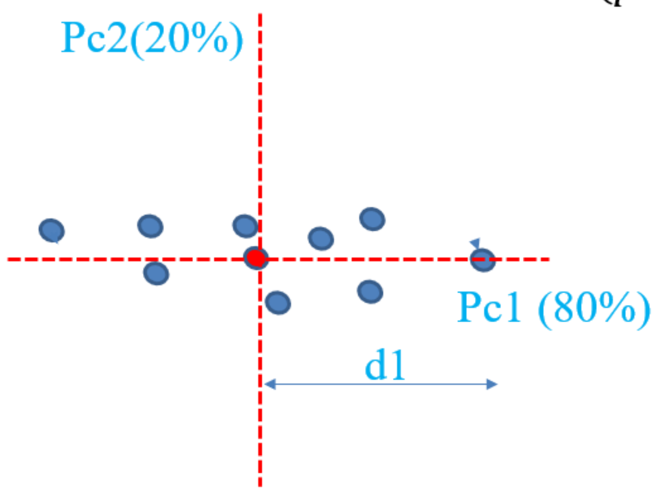
- Poderiamos eliminar X1 e X2 e plotar o gráfico em função de PC1 e PC2.

$$s^2 = \frac{\sum_{i}^{N}(X-\overline{X})^2}{N-1}$$

$$s^2(pc1) = \frac{d1^2 + d2^2 \ldots}{n-1} = 20$$

$$s^2(pc2) = 5$$





INSTITUTO FEDERAL
Sudeste de Minas Gerais

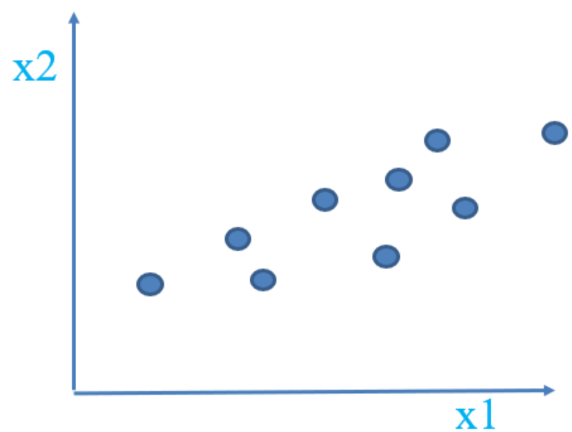
PCA - Análise de Componentes Principais

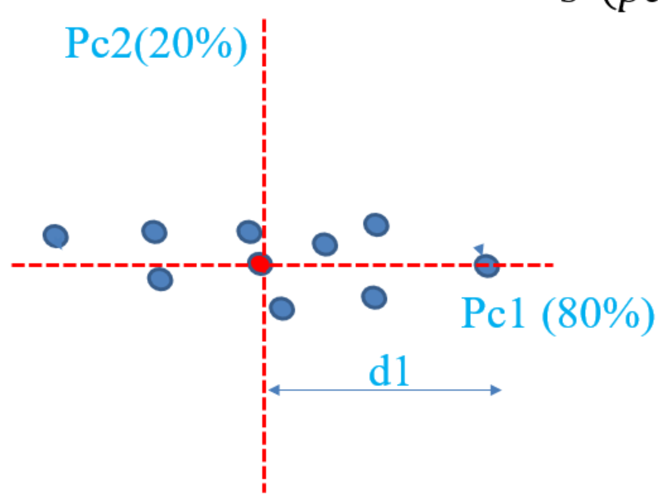
- Veja que 80% da variabilidade dos dados estão sendo explicados pela PC1 e 20% pela PC2.
- Para calcular essa variabilidade utilizando a variância (Est. I).
- Como a média é zero (ponto vermelho), poderíamos eliminar $\overline{X}$

$$s^2 = \frac{\sum_{i}^{N}(X - \overline{X})^2}{N-1}$$

$$s^2(pc1) = \frac{d1^2 + d2^2 \ldots}{n-1} = 20$$

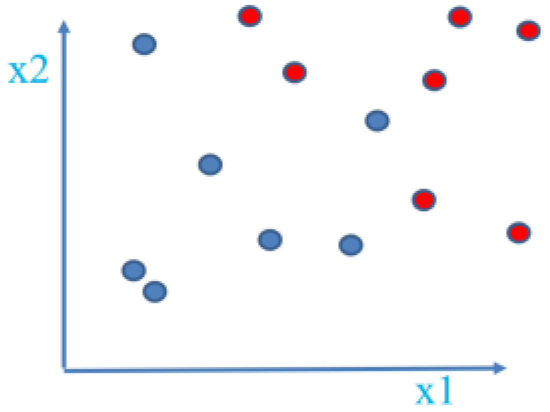
$$s^2(pc2) = 5$$

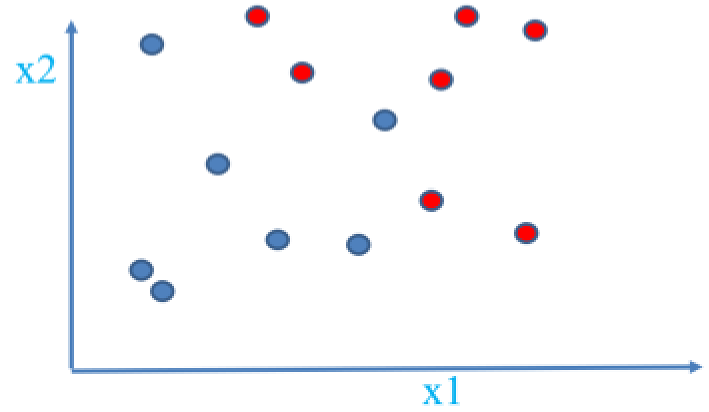




LDA – Linear Descriminante Análise

- Semelhante a PCA, entretanto na LDA temos conhecimento das classes (Y), enquanto que na PCA não temos conhecimento das classes. Então a PCA é não supervisionado enquanto que a LDA e supervisionado.

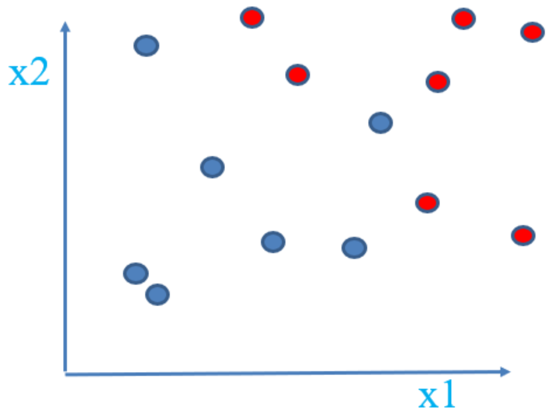


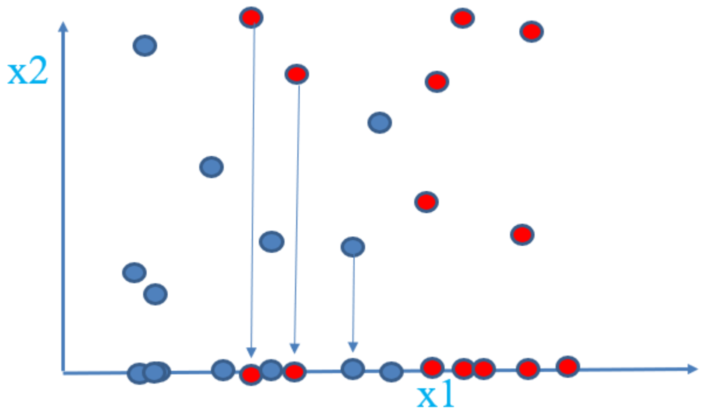


INSTITUTO FEDERAL
Sudeste de Minas Gerais

LDA – Linear Descriminante Análise

- Projetando todos os pontos na X1 teríamos.



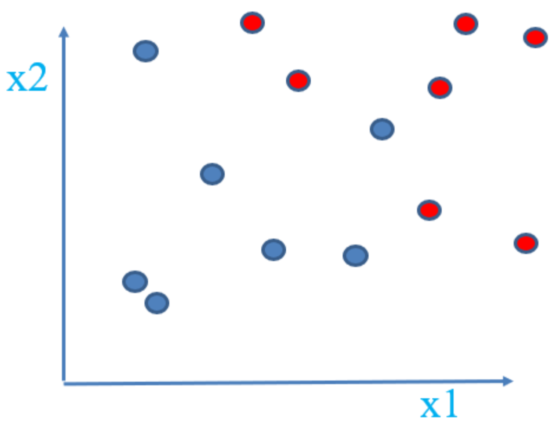


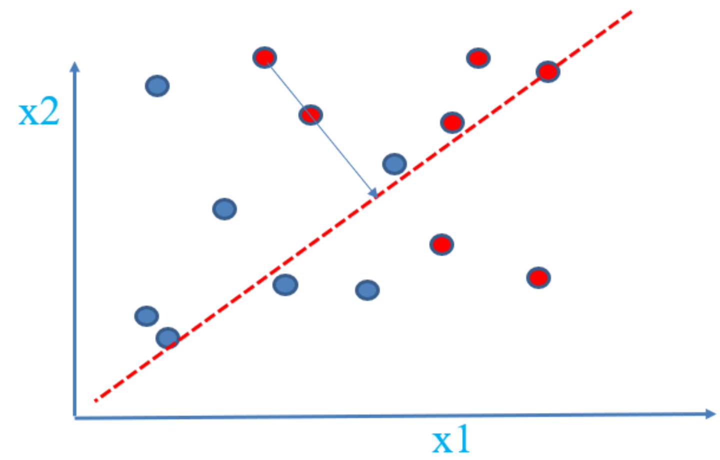
- E se quiséssemos separar (descriminar) estes dados ? Separando as classes vermelhas das azuis.
- Escolhendo um liminar, teríamos uma taxa de erro, pois não seria possível separar todas as vermelhas das azuis.

INSTITUTO FEDERAL
Sudeste de Minas Gerais

LDA – Linear Descriminante Análise

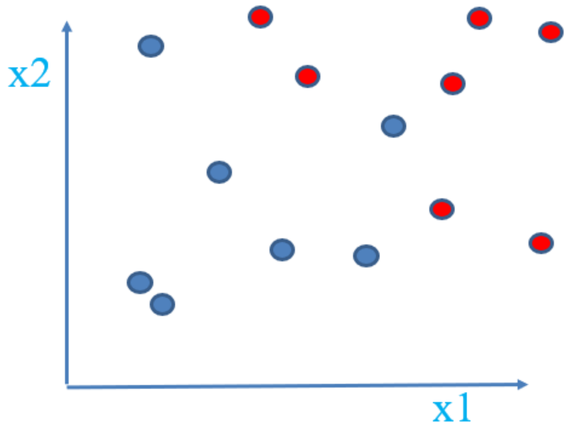
- Mas se fizéssemos diferente, traçando um modelo linear (combinação linear) como o objetivo de maximizar a descriminação das classes ?

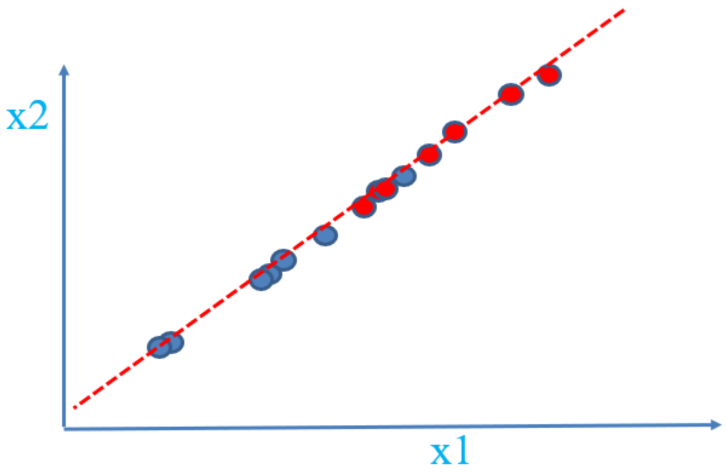




LDA – Linear Descriminante Análise

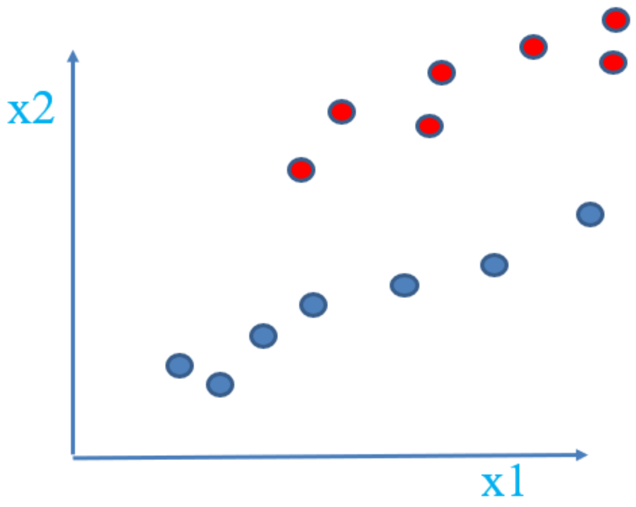
- Veja que agora teríamos apenas um erro ao descriminar as classes.

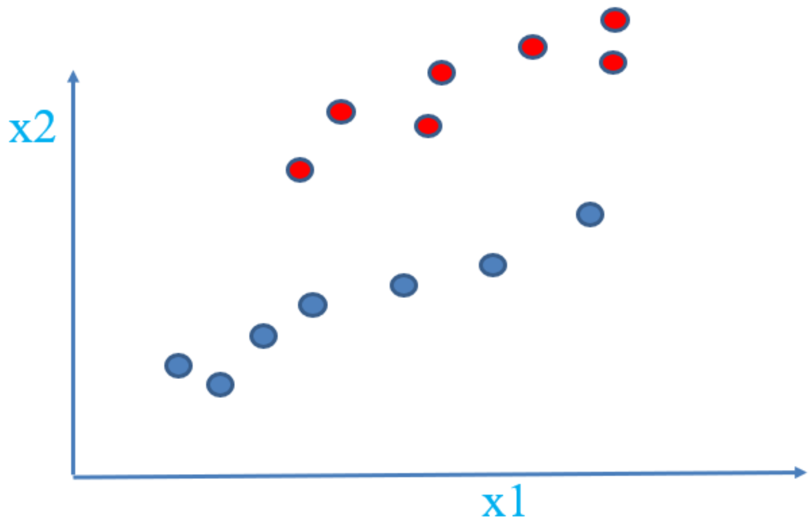




LDA – Linear Descriminante Análise

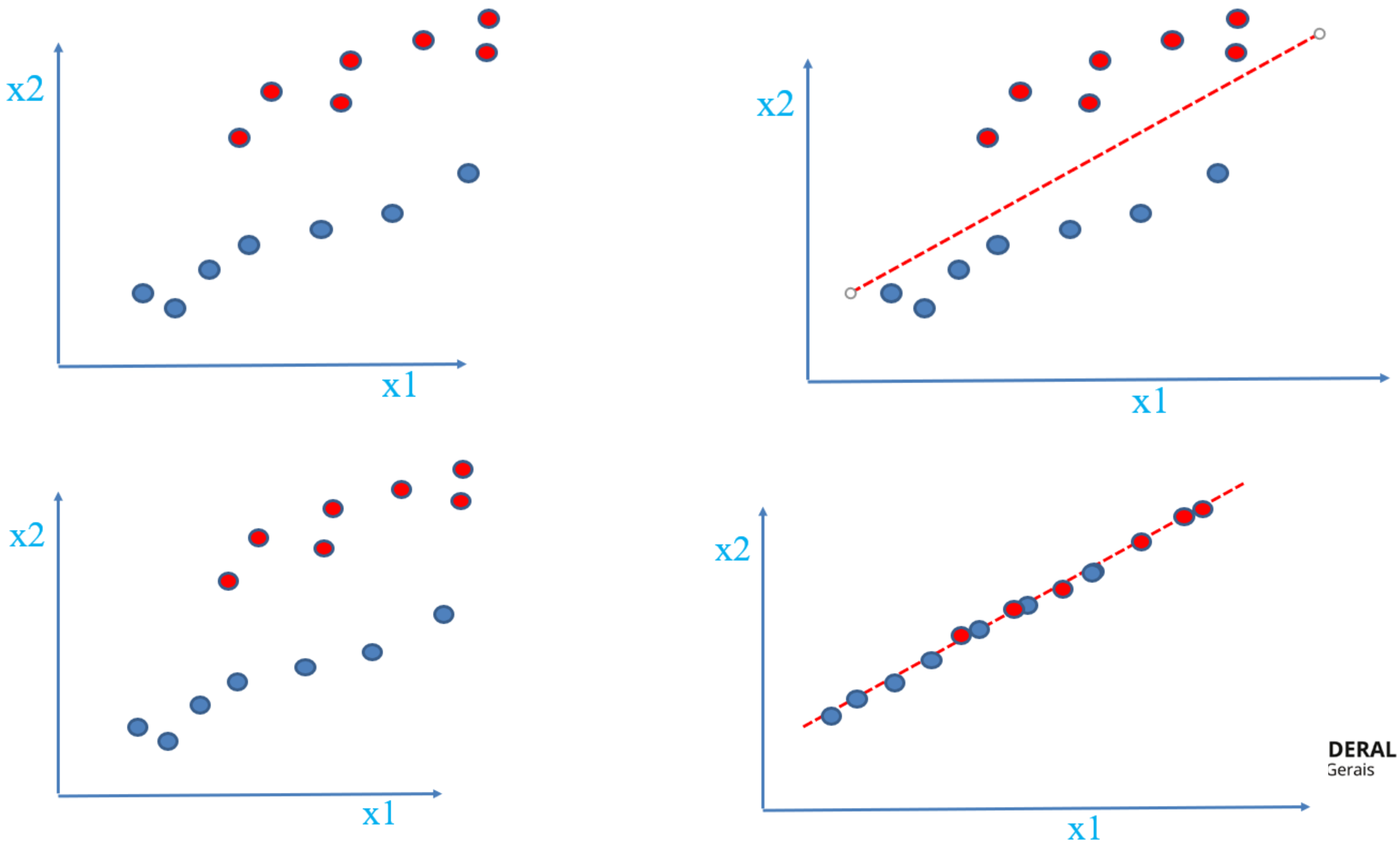
- Agora veja esses outro modelo.
- Como você definiria um modelo para descriminar as classes ?





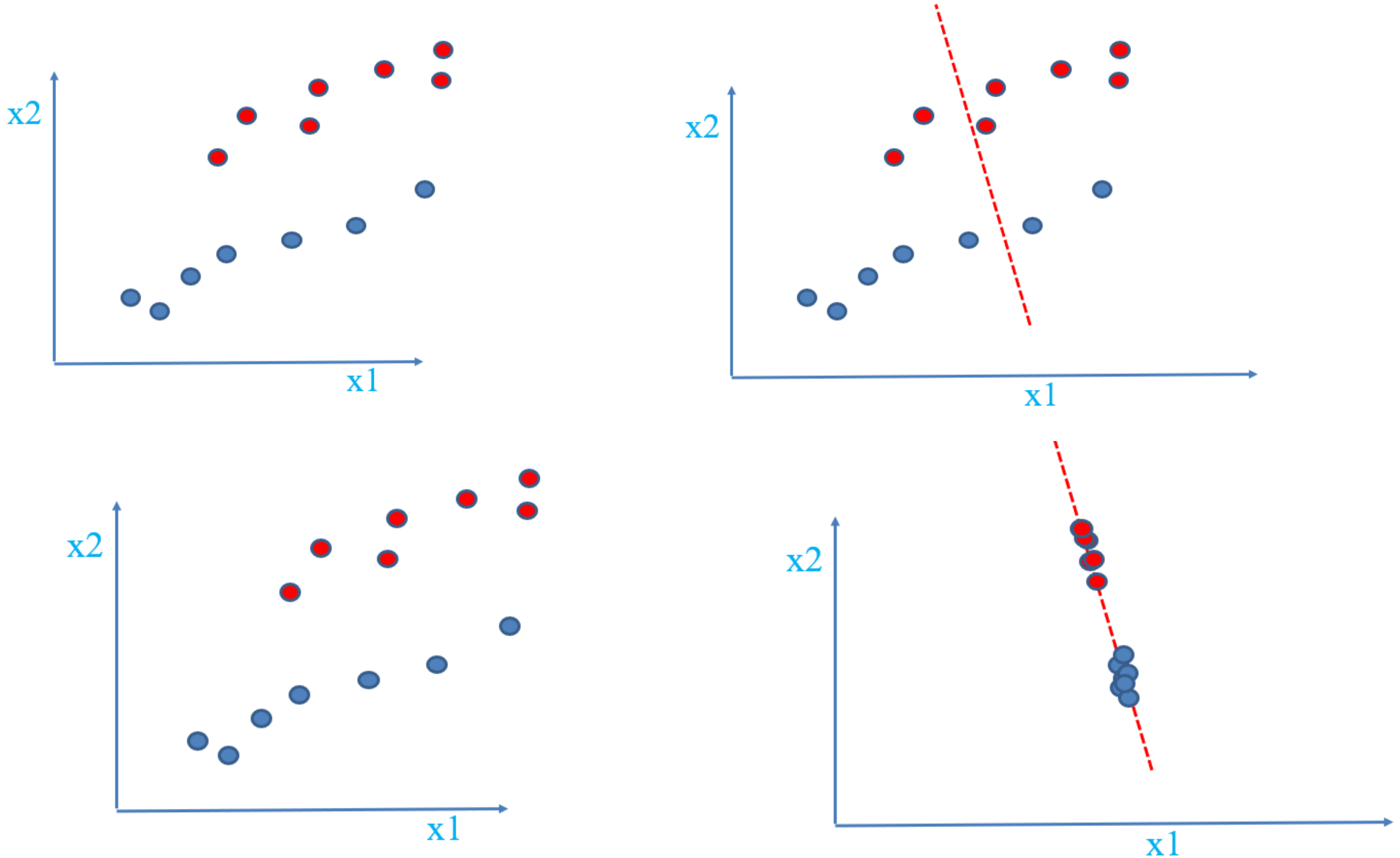
LDA – Linear Descriminante Análise

- Traçando uma reta nessa direção



LDA – Linear Descriminante Análise

- Traçando uma reta nessa outra direção



PCA e LDA - Conclusões

- A PCA tem como objetivo gerar uma combinação linear entre as variáveis que vá em direção a máxima variância dos dados
- A LDA tem como objetivo gerar uma combinação linear entre as variáveis que permite ter a máxima separação (descriminar) as classes.
- Vamos agora ver um código Python para PCA e LDA
- Considere o dataset - framingham descrito nas aulas anteriores onde a variável Y é prever risco de doença cardiovascular futura (DCF).

INSTITUTO FEDERAL
Sudeste de Minas Gerais