

## Avaliação 2 - Individual

1. Dado o problema de classificação de spam, assumo que uma base de dados possua 1500 objetos, sendo 32 o total de spams. O resultado de um algoritmo de classificação aplicado a esta base produziu a seguinte matriz de contingência:

Classe Correta	Classe Predita	
	Spam (positiva)	Normal (negativa)
Spam (positiva)	27 (TP)	5 (FN)
Normal (negativa)	57 (FP)	1411 (TN)

Matriz de Confusão (Spam)

- (1pt) Quantos spams foram classificados corretamente?
  - (1pt) Quantos e-mails não spam foram classificados corretamente?
  - (1pt) Obtenha valores de TPR, FPR, ACC, E e P
2. (7pts) Obtenha uma base de dados no site da UCI (<https://archive.ics.uci.edu/ml/index.php>) que possua classificação binária (ex.: Statlog Heart).
- Faça a normalização dos dados
  - Separe 10% dos dados (5% de cada classe) para considerarmos como dados de teste
  - Executar o kNN para prever os dados de teste; considere os valores de k: 1, 3, 5
  - Faça três funções em Octave para se obter os seguintes dados:
    - Uma matriz de confusão
    - O cálculo de TPR, de FPR e de ACC
    - O gráfico da curva ROC. Use a função `axis([0,1,0,1])` após a função `plot`.
- Faça um *script*, denominado `Principal.m`, o qual deverá carregar os objetos da base e de teste e seus rótulos, os valores de k, e chamar as funções criadas acima
- Obtenha a matriz de confusão para cada valor de k. Obtenha também os respectivos valores para TPR, FPR e ACC.
  - Obtenha a curva ROC para a comparação de resultados obtidos pelos valores de k.
  - Qual valor de k você considera mais interessante para este problema e por quê?

### Entregar:

Códigos e único relatório para as Questões 1 e 2 (entrega em PDF)