

# An SVM classifier for predicting Hypertention in patients

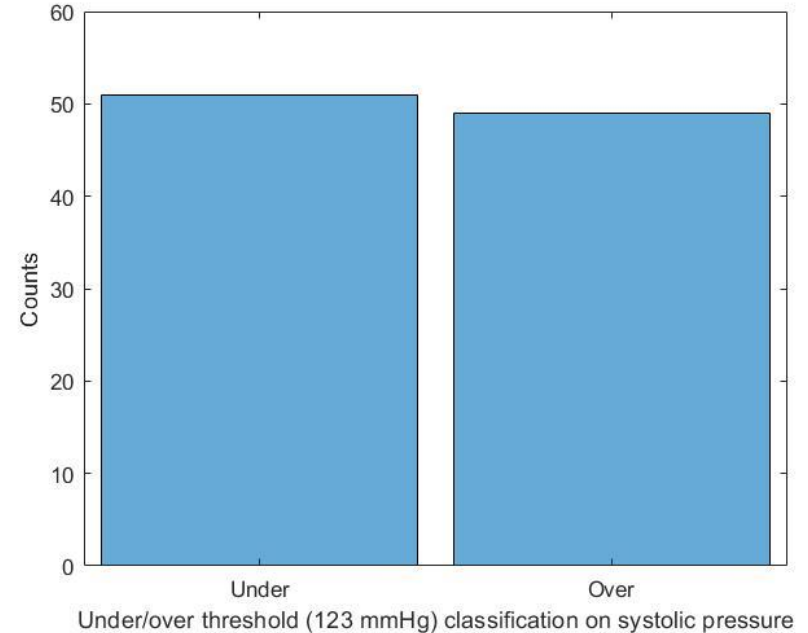
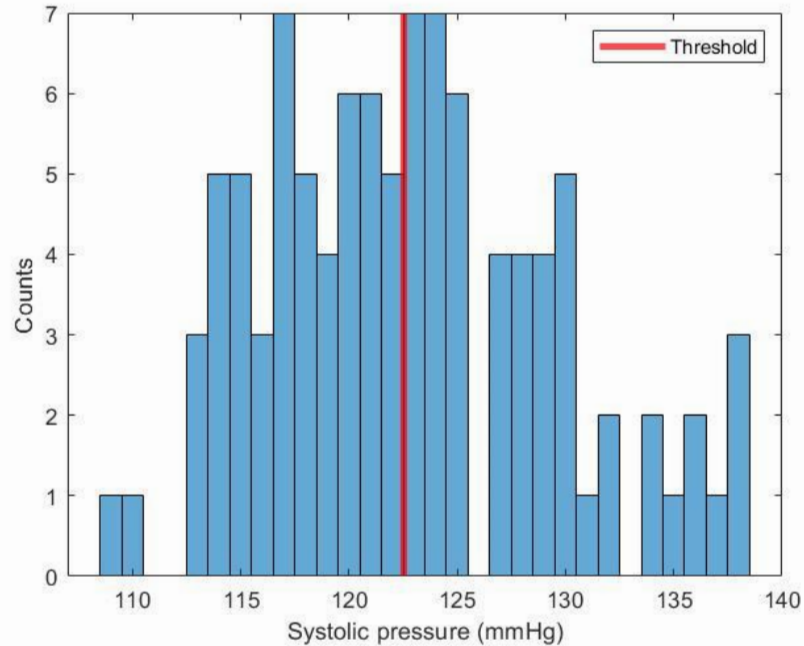
Autori: Comotti F. Covali A. di Noia C. Rossini R. Zappa C.

# Introduzione

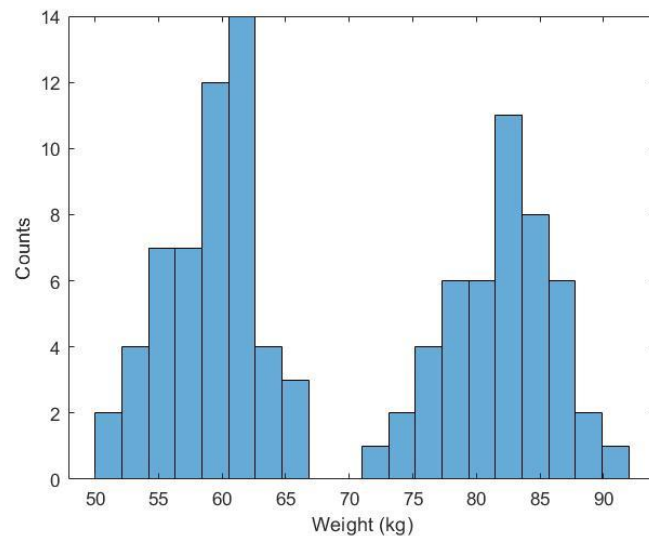
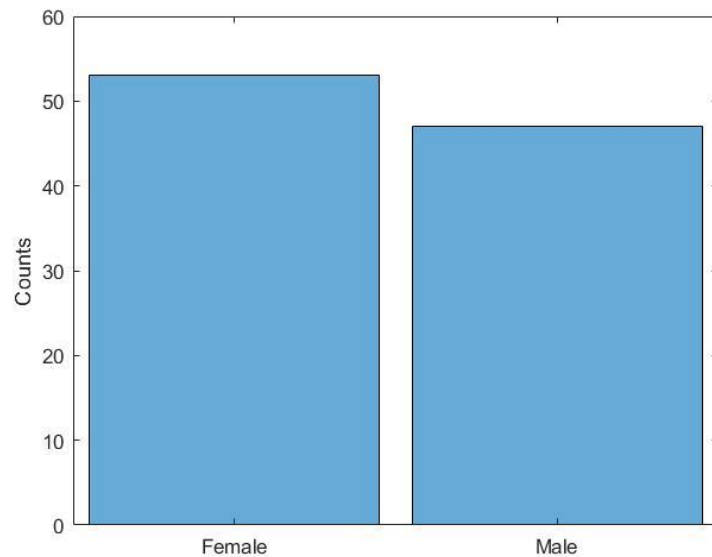
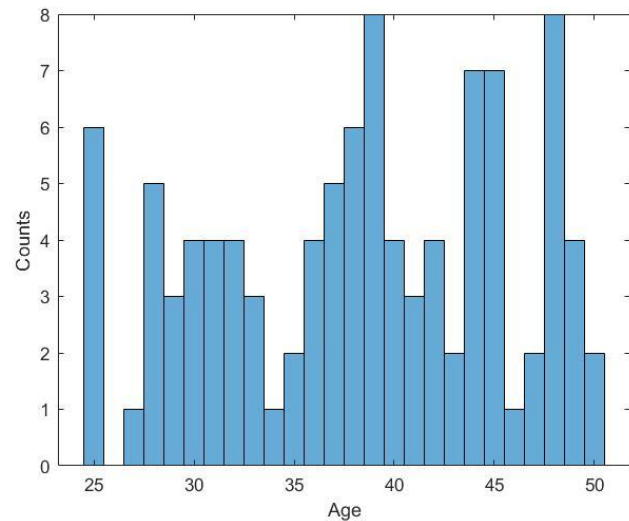
- Dataset di 100 campioni con 5 Features di classificazione (Sesso , Peso, Età , Fumatore e Pressione Diastolica) e una variabile di risposta (pressione Sistolica).
- Classificazione binaria con una soglia è stata fissata a 123 mmHg.
- Features scelte in base al possibile impatto sulla pressione del paziente.
- Utilizzo di 5-fold-CV.
- Algoritmo utilizzato SVM-kernel Linear.

# Variabile di risposta

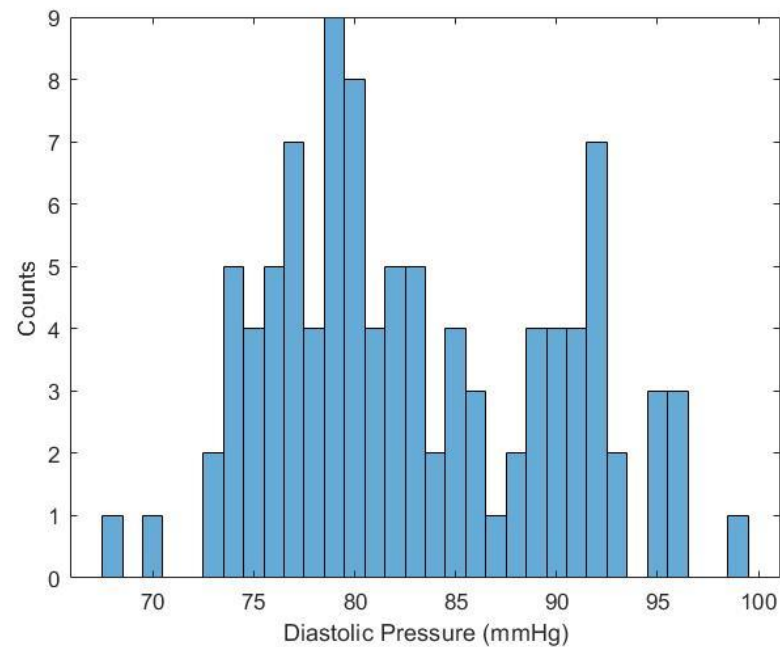
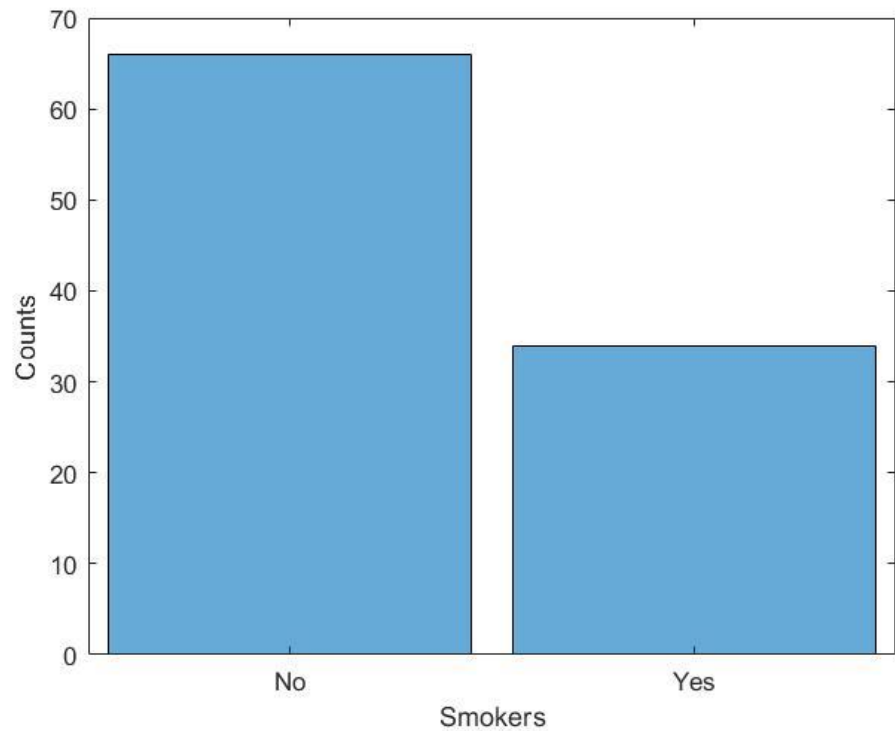
- Classificazione dei pazienti nella classe di risposta definita dalla soglia di 123mmHg




# Features



# Features

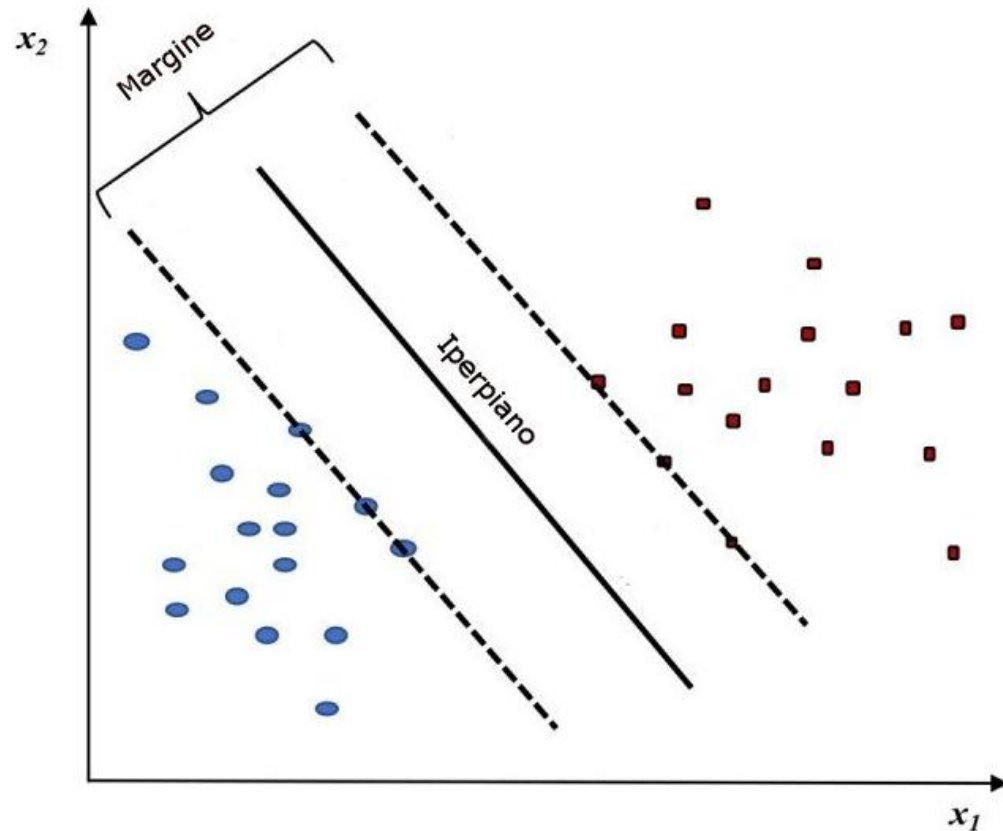


# K-fold Cross-Validation

- Cross-Validation procedura utilizzata per valutare i modelli di machine learning su un campione di dati limitati.
- Parametro **K** = numero di gruppi in cui viene suddiviso il campione di dati.
- La CV restituisce una stima meno distorta, dell'abilità del modello, rispetto ad altre procedure.
- La procedura è la seguente:
  - Mescola il set di dati in modo casuale,
  - Divide il set di dati in K gruppi,
  - Prende un gruppo come set di dati di prova,
  - Prende i K-1 gruppi restanti come set di dati di addestramento  valutato su set di prova
  - Conserva punteggio ed elimina modello
  - Riassume le abilità del modello grazie al campione di punteggi di valutazione.

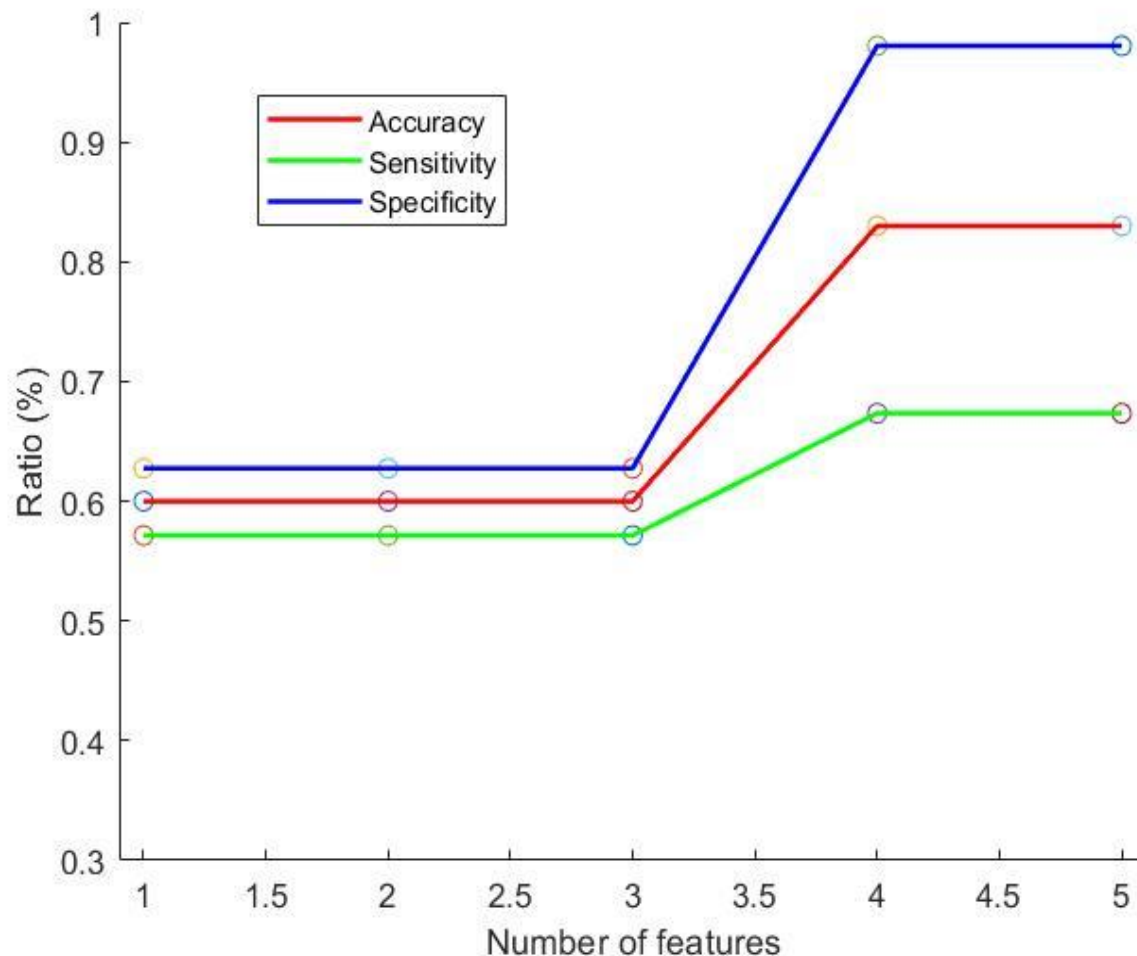
# Svm (Support Vector Machine)

- Due classi di pattern linearmente separabili.
- Training set contenente 80 campioni.
- L'algoritmo si ripete 5 volte per ogni features aggiunta.



# Risultati

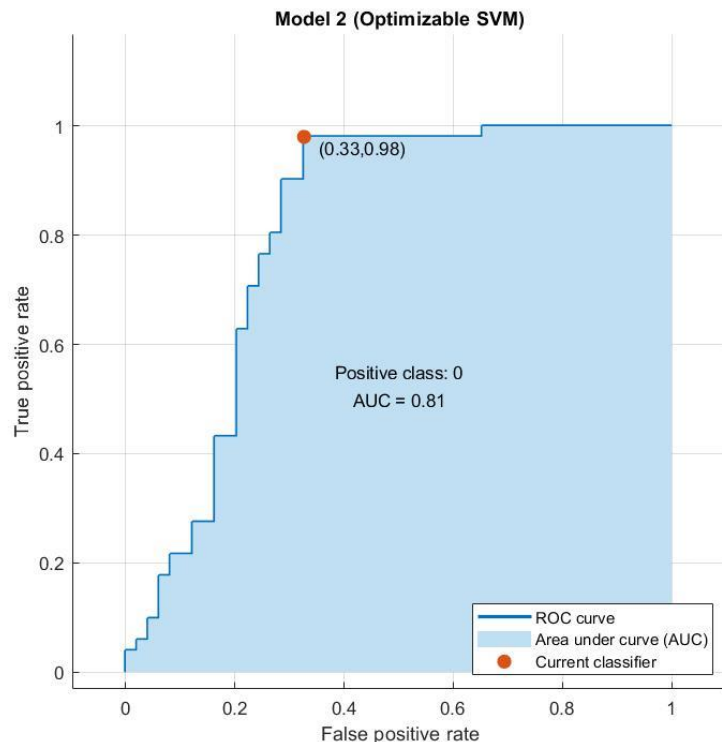
- Come si può notare i risultati della accuracy di classificazione rimane invariata con le features (sesso, età, peso) per poi avere un netto incremento con l'utilizzo della feature Fumatori.
- Si può concludere quindi che la Fumatori possiede un netto impatto nel task di classificazione.





# Tool Classification Learner

- Abbiamo infine utilizzato un tool di Matlab per confrontare con altri modelli di classificazione come il KNN e lo stesso SVM gli stessi risultati.



## Model 2: Trained

### Training Results

Accuracy (Validation)	83.0%
Total cost (Validation)	17
Prediction speed	~2700 obs/sec
Training time	251.03 sec

### Model Type

Preset: Optimizable SVM  
Kernel scale: 1

### Optimized Hyperparameters

Kernel function: Linear  
Box constraint level: 0.059235  
Multiclass method: One-vs-All  
Standardize data: true

### Hyperparameter Search Range

Multiclass method: One-vs-All, One-vs-One  
Box constraint level: 0.001-1000  
Kernel scale: 0.001-1000  
Kernel function: Gaussian, Linear, Quadratic, Cubic  
Standardize data: true, false

### Optimizer Options

Optimizer: Bayesian optimization  
Acquisition function: Expected improvement per second plus  
Iterations: 30  
Training time limit: false

### Feature Selection

All features used in the model, before PCA

### PCA

PCA disabled

### Misclassification Costs

Cost matrix: default