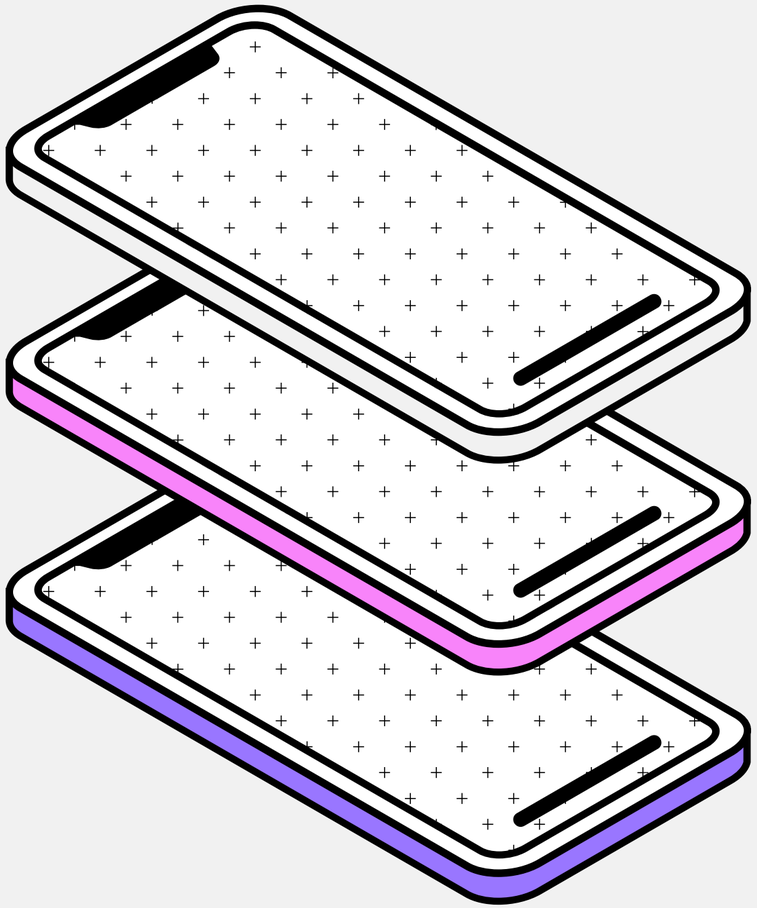
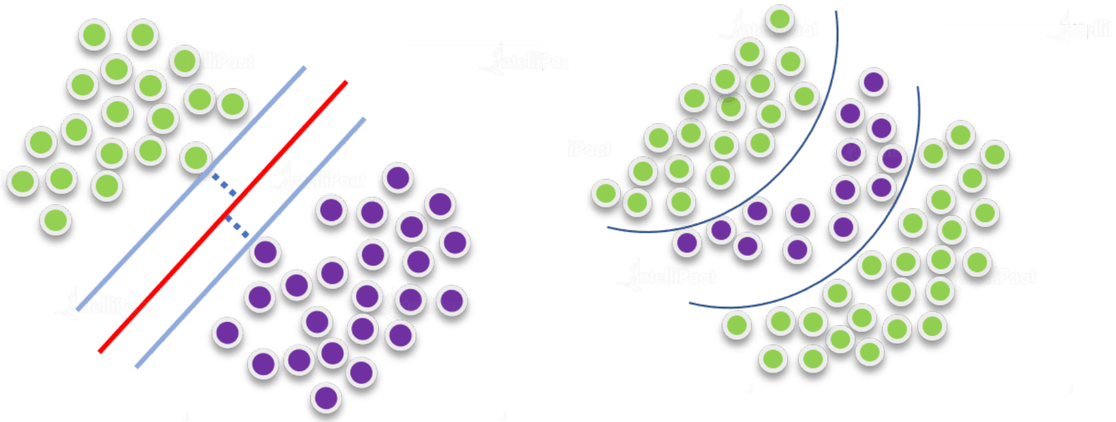
mobile price classification

FERGUOUS Wafa



what is SVM ?



supervised learning algorithm

List the topics covered in the presentation

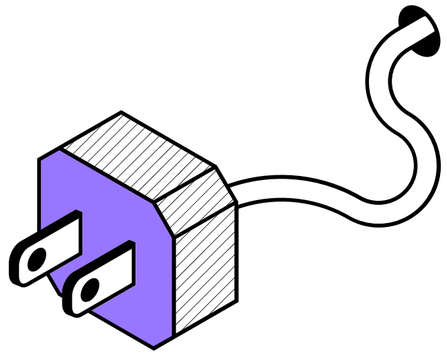
classification and regression

find a hyperplane

useful in cases where the data is not linearly separable

the kernel trick

the problem?



Mr Bob wants to start his own mobile phone company and he wants to wage an uphill battle with big smartphone brands like Samsung and Apple. But he doesn’t know how to estimate the price of a mobile that can cover both marketing and manufacturing costs. So in this task, we don’t have to predict the actual prices of the mobiles but we have to predict the price range of the mobiles.

about the data

data contains 21 attributes

The class label, price range, is of ordinal data type with 4 values

The 20 features include battery power, bluetooth availability, clock speed, and others

sourced from Kaggle

The problem can be approached with algorithms such as SVM

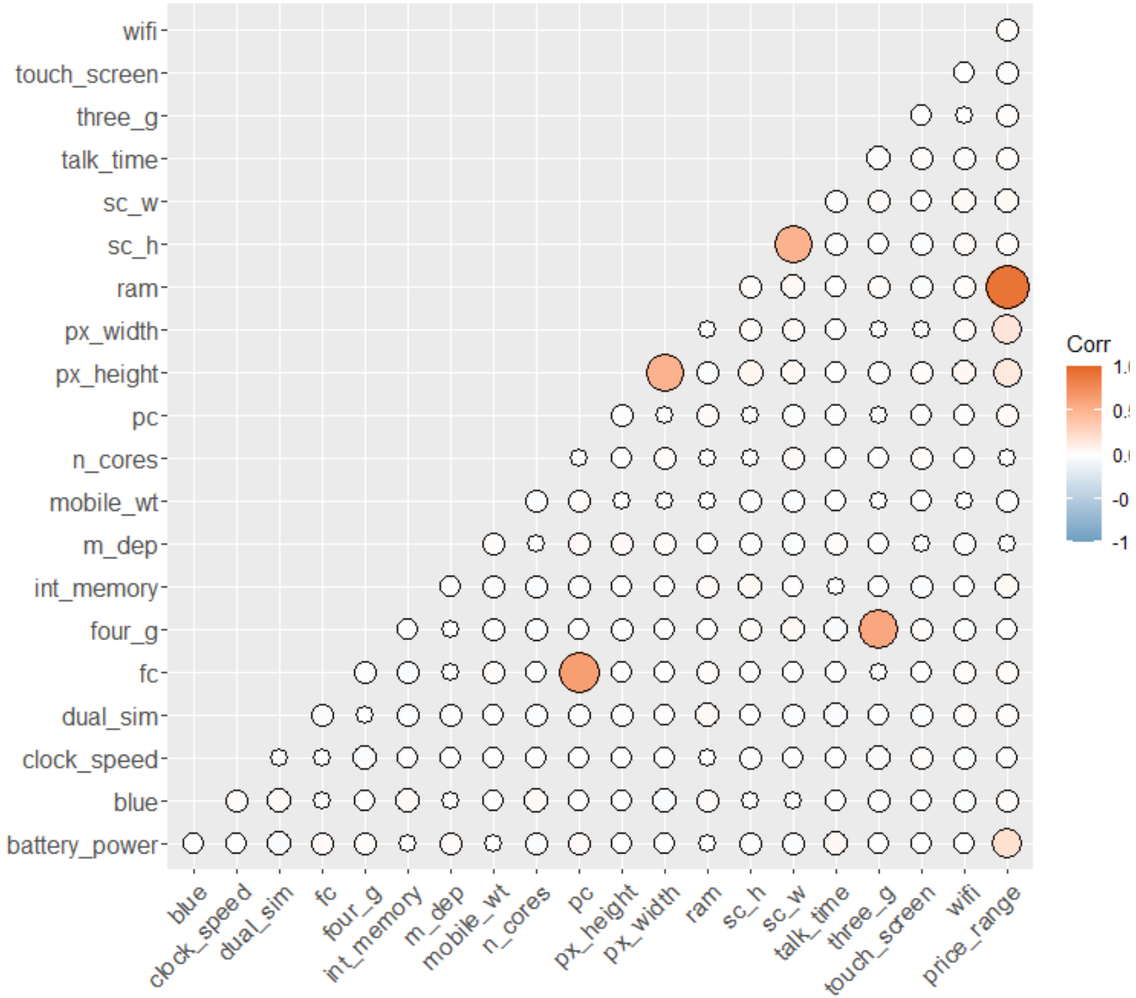
correlation between variables

the highest correlation is found in RAM

Corr(pc, fc) = 0.6

Corr(three_g, four_g) = 0.6

Corr(px_width, px_heght) = 0.5



what is cross validation ?

Cross-validation is a technique that is used to evaluate the performance of a model by training it on different subsets of the data and testing it on the remaining data. It is a valuable technique when working with SVMs as it can provide an unbiased estimate of the model's performance. It can also help to identify any overfitting that may be present in the model

tuning parameters

Regularization parameter (C)

Kernel parameters

Gamma

Tolerance

and Maximum number of iteration

```
Confusion Matrix and Statistics

   predictions
      0   1   2   3
  0 110   1   0   0
  1   1  91   0   0
  2   0   9  87   6
  3   0   0   3  96

Overall Statistics

               Accuracy : 0.9505
                 95% CI : (0.9246, 0.9695)
    No Information Rate : 0.2748
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9339

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            0.9910   0.9010   0.9667   0.9412
Specificity            0.9966   0.9967   0.9522   0.9901
Pos Pred Value         0.9910   0.9891   0.8529   0.9697
Neg Pred Value         0.9966   0.9679   0.9901   0.9803
Prevalence             0.2748   0.2500   0.2228   0.2525
Detection Rate         0.2723   0.2252   0.2153   0.2376
Detection Prevalence   0.2748   0.2277   0.2525   0.2450
Balanced Accuracy      0.9938   0.9488   0.9594   0.9656
```

```
Confusion Matrix and Statistics

   predictions
      0   1   2   3
  0 108   3   0   0
  1   2  90   0   0
  2   0   6  90   6
  3   0   0   4  95

Overall Statistics

               Accuracy : 0.948
                 95% CI : (0.9216, 0.9675)
    No Information Rate : 0.2723
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9306

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 0 Class: 1 Class: 2 Class: 3
Sensitivity            0.9818   0.9091   0.9574   0.9406
Specificity            0.9898   0.9934   0.9613   0.9868
Pos Pred Value         0.9730   0.9783   0.8824   0.9596
Neg Pred Value         0.9932   0.9712   0.9868   0.9803
Prevalence             0.2723   0.2450   0.2327   0.2500
Detection Rate         0.2673   0.2228   0.2228   0.2351
Detection Prevalence   0.2748   0.2277   0.2525   0.2450
Balanced Accuracy      0.9858   0.9513   0.9594   0.9637
```

thank you !