K-NEAREST NEIGHBORS

Kumar Sundram

k-nearest neighbors

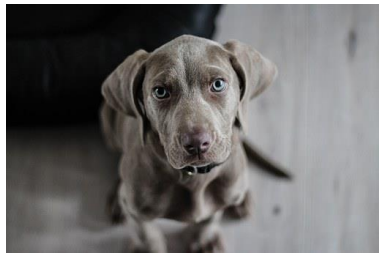
- KNN Intitution
- Compare the two data points, distance metric
- How to choose a K
- Implement KNN from scratch
- Implement KNN on Iris data set with Scikit Learn
- Use KNN as Regression
- Exercise : Apply KNN on MNIST Data

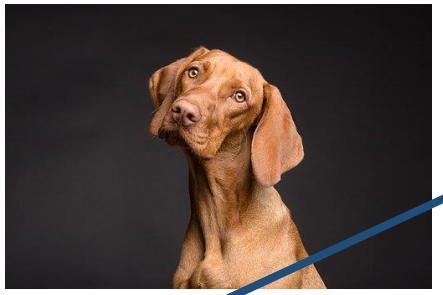
KNN INTITUTION

Problem

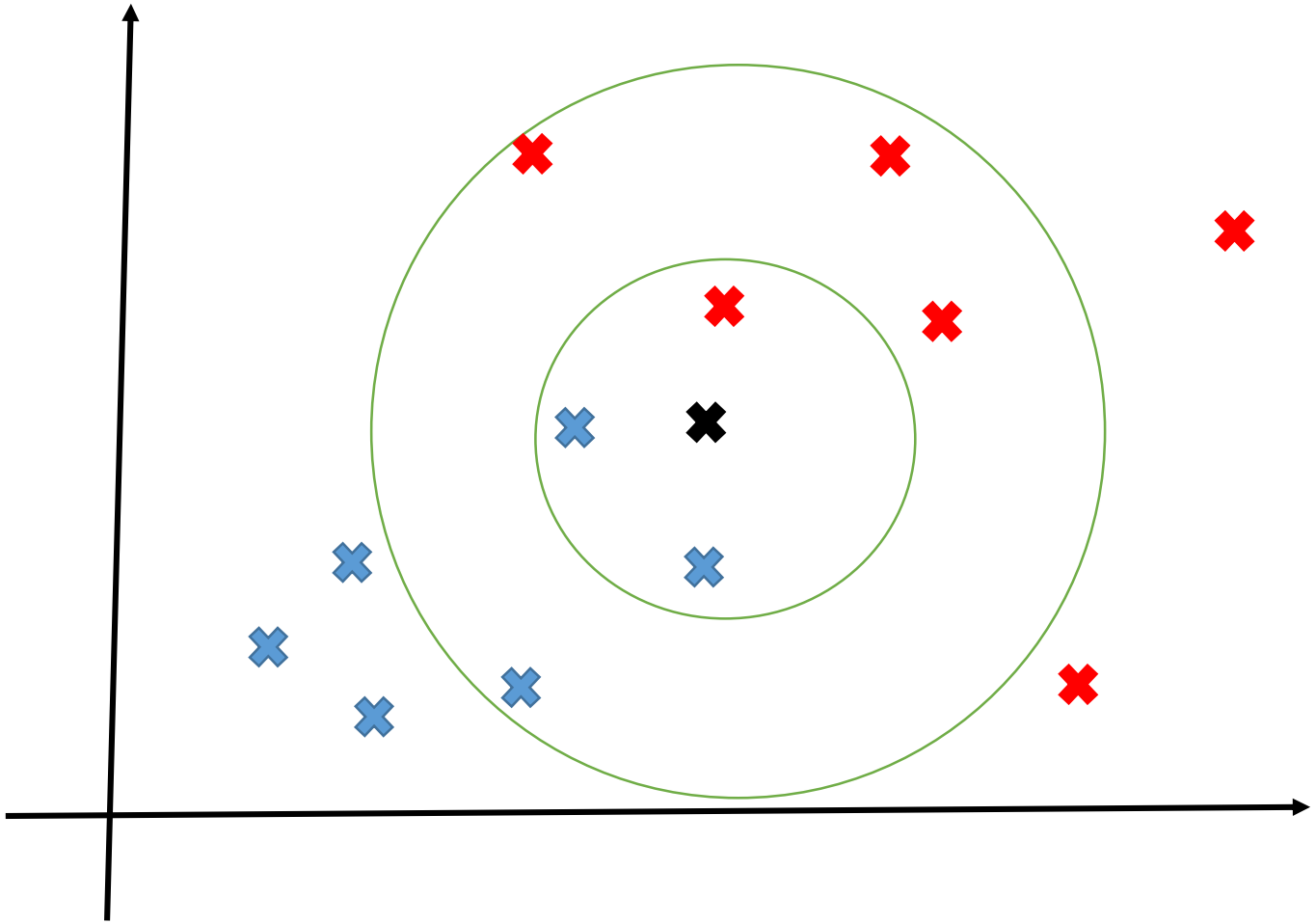
Classification Problem –

Dog or Cat





Visualize KNN



K = 3

K = 5

KNN Algorithm

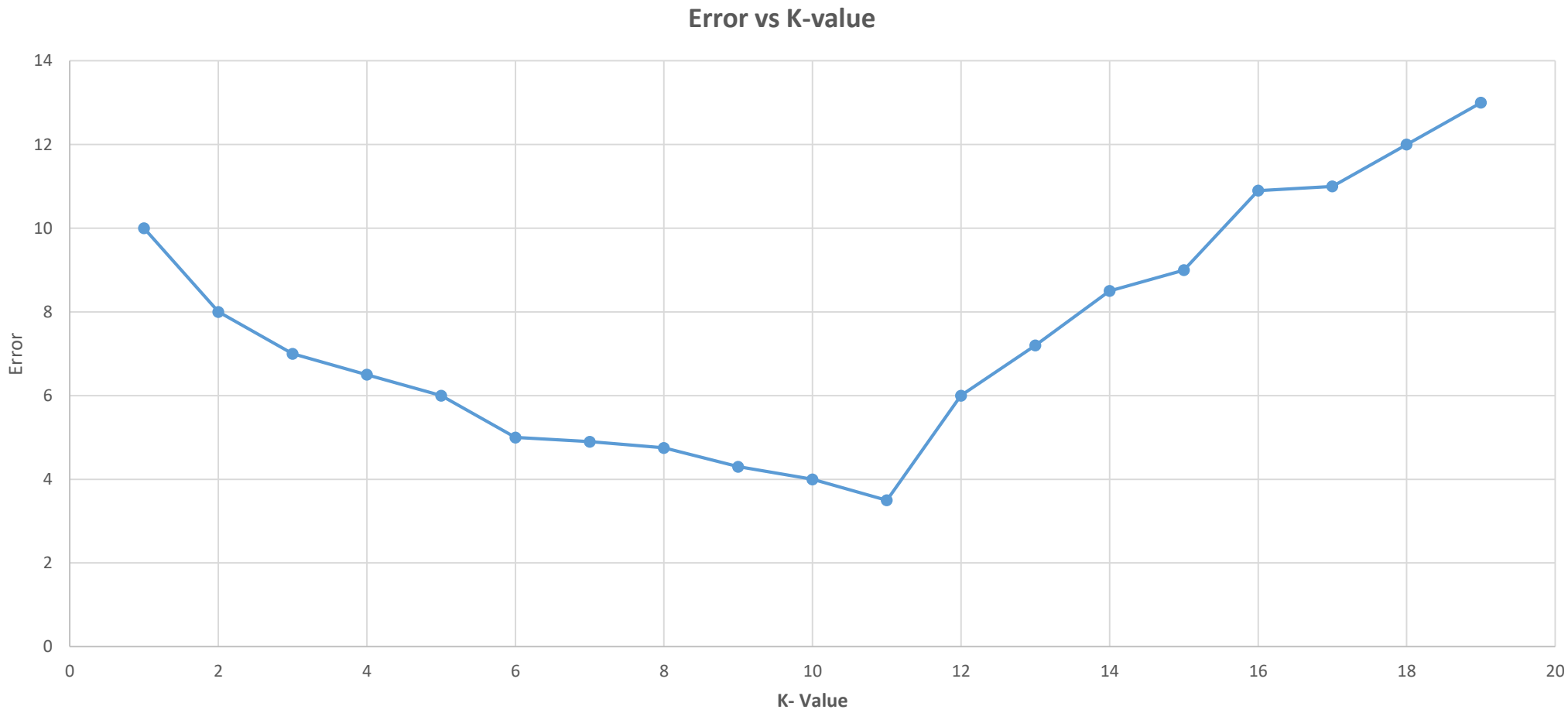
- Choose K
- Find Distance with all point from test point
- Sort Distance
- Select K point with Minimum distance
- Classify in majority vote

Question Arise

- How many K – nearest neighbor
- How to Compare

How many K

Error vs K-value



How to Compare

Distance Measure

$$Euclidien\ Distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

“euclidean”	EuclideanDistance	sqrt(sum((x - y)^2))
“manhattan”	ManhattanDistance	sum(|x - y|)
“chebyshev”	ChebyshevDistance	max(|x - y|)
“minkowski”	MinkowskiDistance	sum(|x - y|^p)^(1/p)
“wminkowski”	WMinkowskiDistance	sum(|w * (x - y)|^p)^(1/p)
“seuclidean”	SEuclideanDistance	sqrt(sum((x - y)^2 / V))
“mahalanobis”	MahalanobisDistance	sqrt((x - y)' V^-1 (x - y))

Source : http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.DistanceMetric.html

Some Aspect of KNN

- Lazy learner
- Instance Based
- Just memorization of Data
- Easy interpretation of result
- Performance on Read world data – not very good
- Computational cost very high for big dataset
- Non-parametric in nature

IMPLEMENT KNN FROM SCRATCH

Steps

- Importing Library and Load Data
- Normalize Data - scaling
- Split data – Train Test
- Apply KNN on 1 Test Sample
- Accuracy