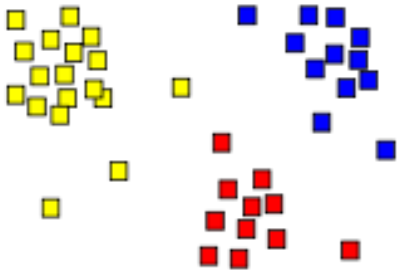
Classification
CS1D6: Introduction to data and statistics

Dr. Fayyaz Minhas

Department of Computer Science
University of Warwick

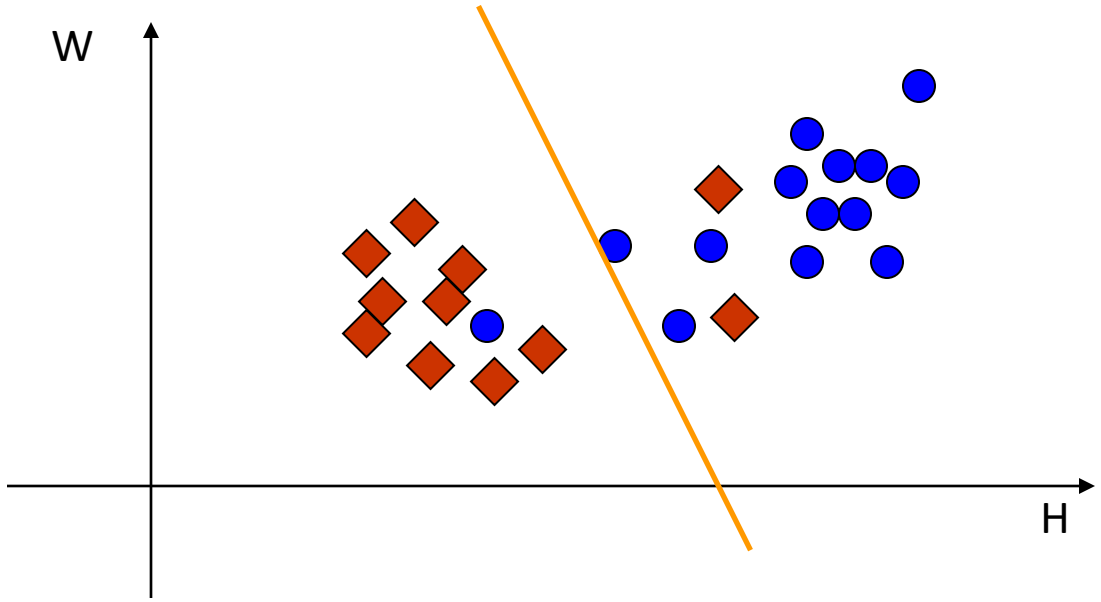
The problem

- How to classify objects?



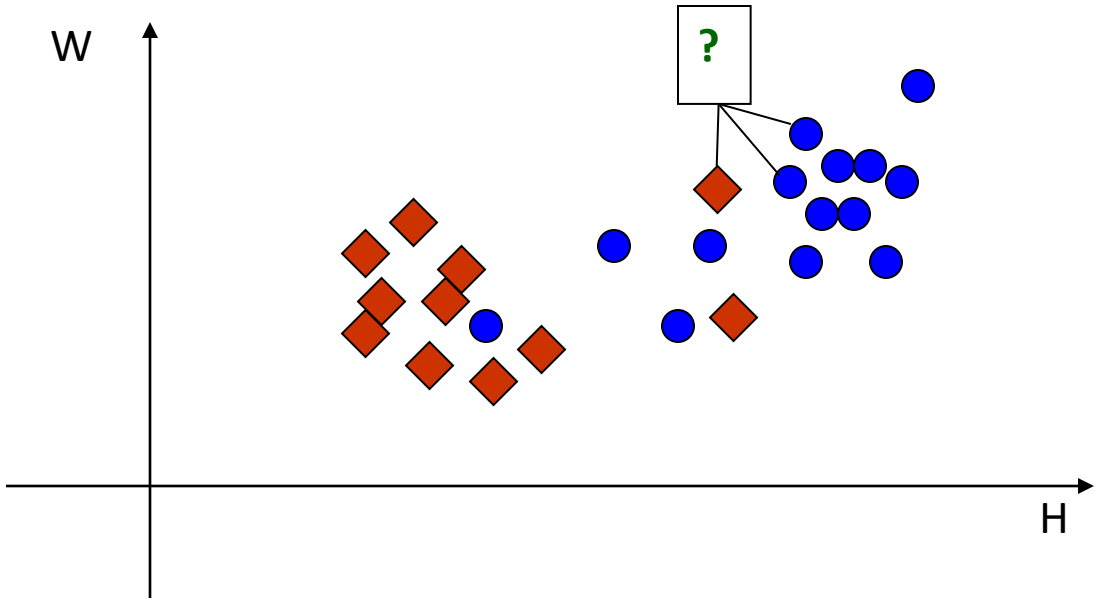
Classification Approaches: Supervised…

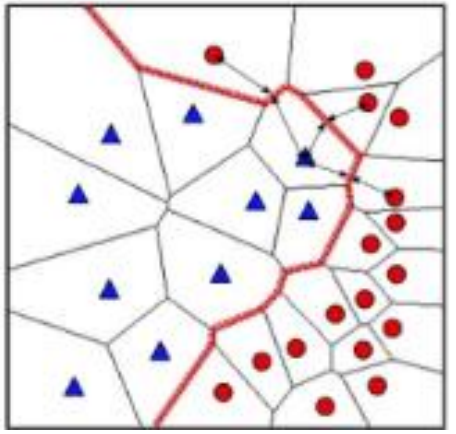
- Linear Classifier



Classification Approaches: Supervised

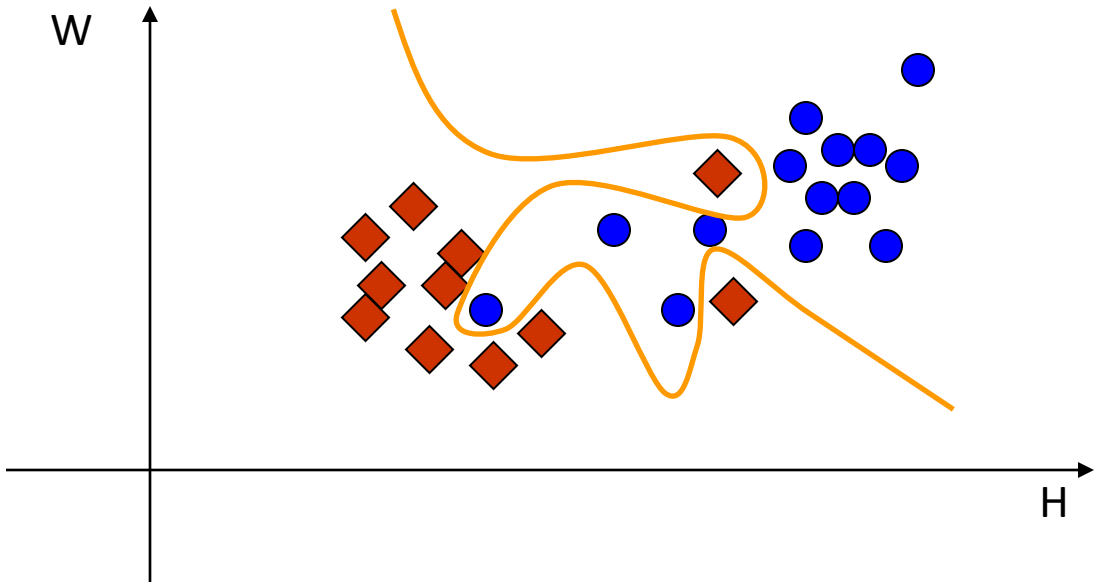
- Example (k=3)-Nearest Neighbor Classification

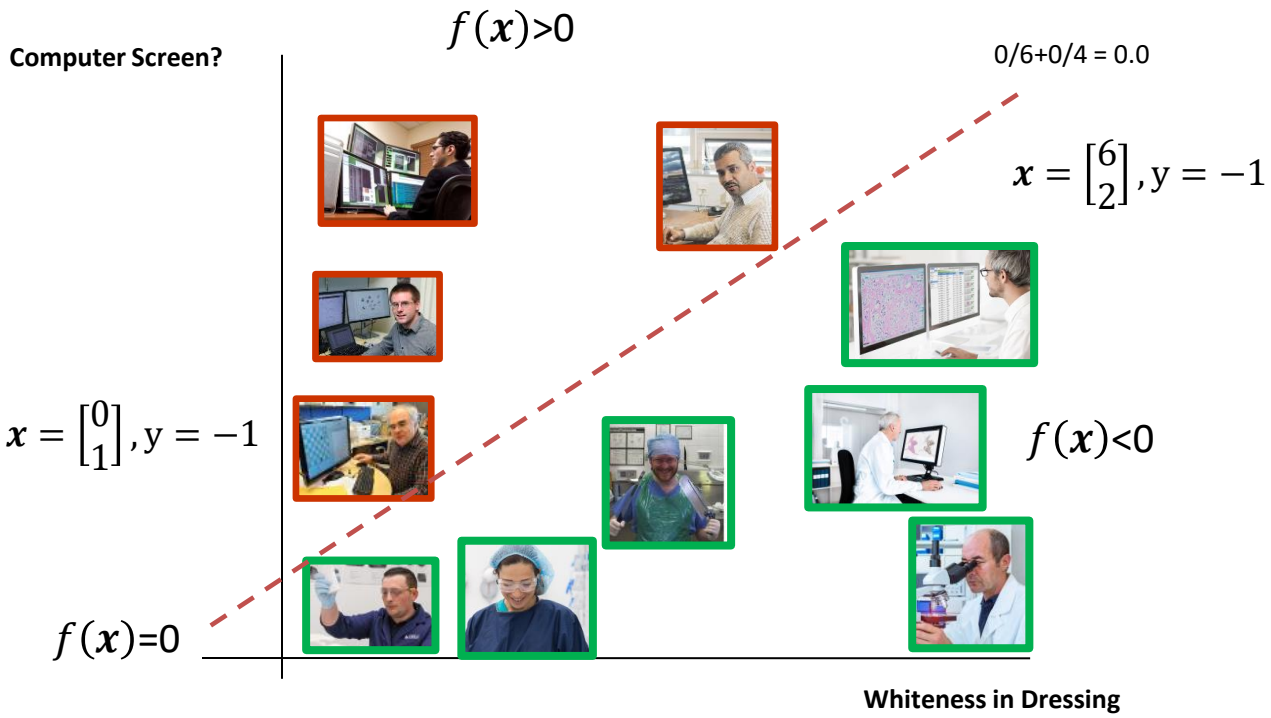




Classification Approaches: Supervised…

- Nonlinear Classification boundary




$f(x)$>0
Computer Screen?
0/6+0/4 = 0.0
$x = \begin{bmatrix} 6 \\ 2 \end{bmatrix}, y = -1$
$x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, y = -1$
$f(x)$<0
$f(x)$=0
Whiteness in Dressing

Classification using k-NN

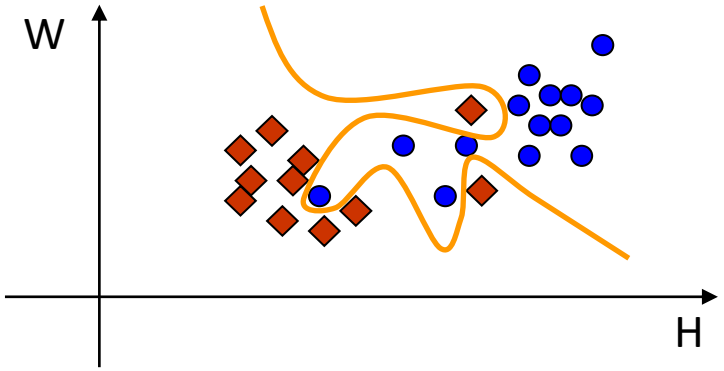
- K-Nearest Neighbor Algorithm
  - Given a dataset $\{(\boldsymbol{x}_i, y_i) | i = 1 \dots N\}$, classify a given test example $\boldsymbol{x}$ to belong to class y based on the majority of the class labels of its K nearest neighbors

Machine Learning Workflow

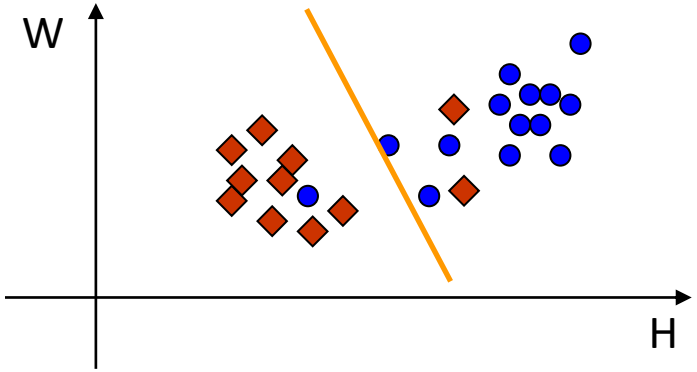
- Training Data
- Feature Extraction
- Training
- Performance Assessment over a validation set
- Deployment

Classification Approaches: Supervised…

- Generalization vs. Memorization
  - A particular issue in classification is the tradeoff between memorization vs. generalization
    - Remembering everything is not learning
    - The true test of learning is handling similar but unseen cases



Has great memorization but may generalize poorly



Has lesser memorization but may generalize better

Example

- Classification in Sk-learn

https://www.kaggle.com/nautna/iris-knn-python-classification

End of Lecture

We want to make a machine that will be proud of us.

- Danny Hillis