EC-350 AI and Decision Support Systems

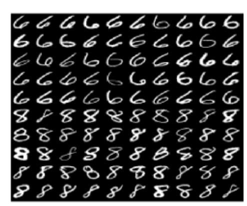
Week 8
Introduction to Machine Learning

Dr. Arslan Shaukat

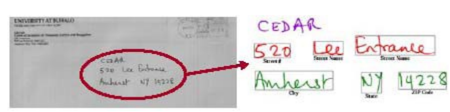
NUST

Machine Learning Applications

- Optical Character Recognition (OCR)
  - *Sorting letters by postal code*
  - *Reconstructing text from printed materials*
  - *Handwritten character recognition*

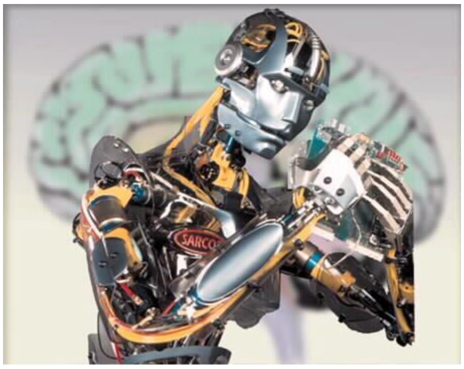


- Analysis and identification of human patterns (Biometric classification)

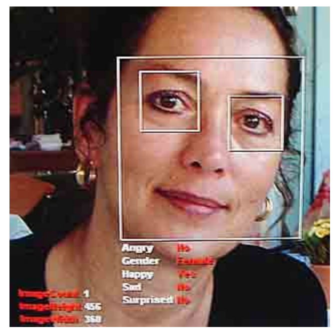


  - *Face recognition*
  - *Handwriting recognition*
  - *Fingerprints and DNA sequence identification*
  - *Iris scan identification*
  - *Speech recognition/speaker identification*

Applications

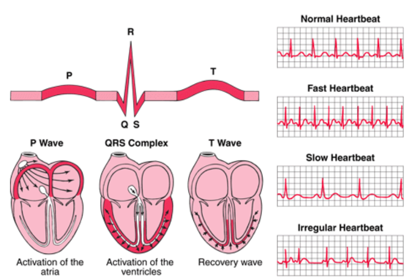


amazon.com



Examples of ML

- Computer aided diagnosis
  - *Medical imaging, EEG, ECG signal analysis*
  - *Designed to assist (not replace) physicians*
- Prediction systems
  - *Weather forecasting (based on satellite data)*
- Information Retrieval
  - *Data Mining*



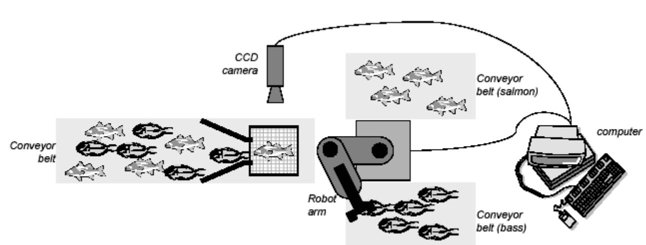
Machine Perception and Pattern Recognition

- Machine Perception
  - *Build a machine that can recognize patterns*
- Pattern Recognition
  - *Theory, Algorithms, Systems to Put Patterns into Categories*
  - *Relate Perceived Pattern to Previously Perceived Patterns*

- By building such systems, we gain understanding of machine learning, particularly in humans

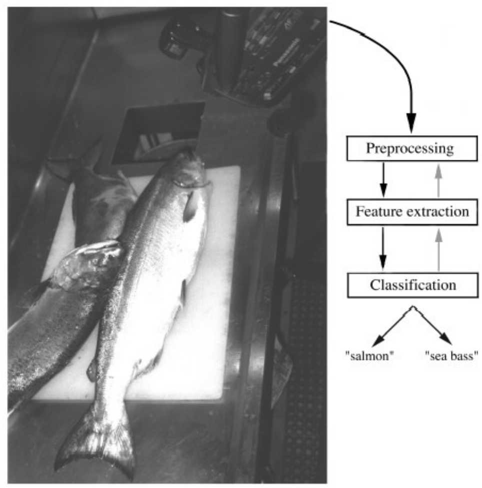
A Machine Learning Example

- A fish processing plant wants to automate the process of sorting incoming fish according to species (salmon or sea bass)
- The automation system consists of
  - *A conveyor belt for incoming products*
  - *A vision system with an overhead camera*
  - *A computer to analyze images and control the robot arm*



Example

- "Sorting incoming fish on a conveyor belt according to species using optical sensing".



Problem Analysis

- Set up a camera and take some training sample images to extract features.
  – *Length*
  – *Lightness*
  – *Width*
  – *Number and shape of fins*
  – *Position of the mouth, etc…*
- This is the set of all suggested features to explore for use in our classifier
- Purpose:
  – *To classify the future samples based on the data of extracted features from the training samples*

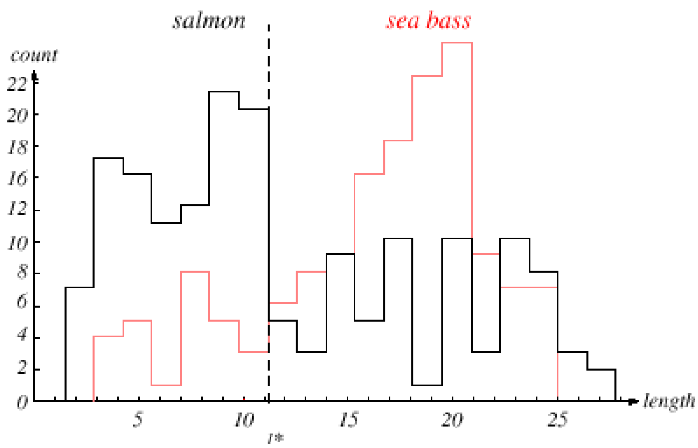
Example

- Models
  - *There are differences between sea bass and salmon and are viewed as having different models.*
- Preprocessing
  - *Segmentation*
  - *Isolate fish from one another and from the background.*
- Feature Extraction
  - *Information from a single fish is sent to a feature extractor whose purpose is to reduce the data by measuring certain features.*
- Classification
  - *Evaluates the evidence presented and makes a final decision.*

Selection Criterion

- Suppose sea bass is generally longer than a salmon.
- Select only the length of the fish as a possible feature for discrimination.
- To choose critical value of length, we could obtain some design or training samples of the different types of fish.

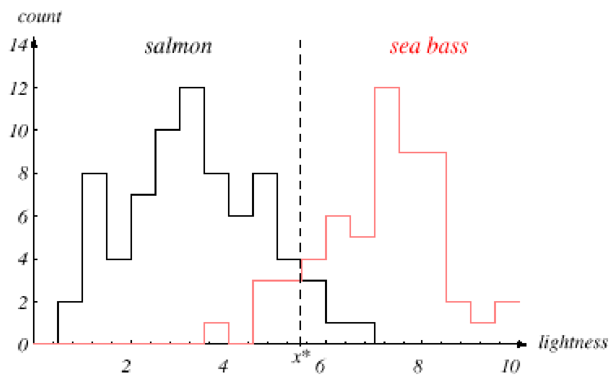
Histograms for the Length Feature



Selection Criterion

- No matter how we choose the threshold value of length, we cannot reliably separate sea bass from salmon.
- The length is a poor feature alone!
- Select the average lightness of the fish scales as a possible feature.

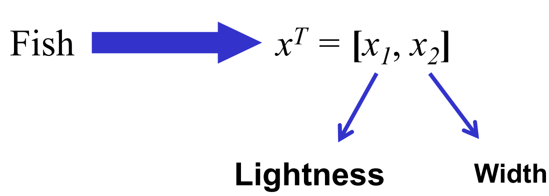
Histograms for the Lightness Feature



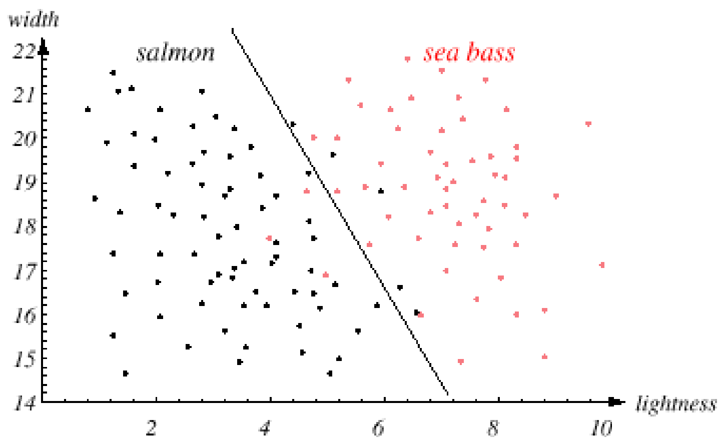
- Length or Lightness, which one is a better feature?
- No value of either feature will “classify” all fish correctly

Selection Criterion and Decision Boundary

- Seek a different feature to separate the fish.
- Use more than 1 feature at a time.
- Adopt the lightness feature.
- Add the width of the fish.
- *Feature vector x* is a 2D *feature space.*



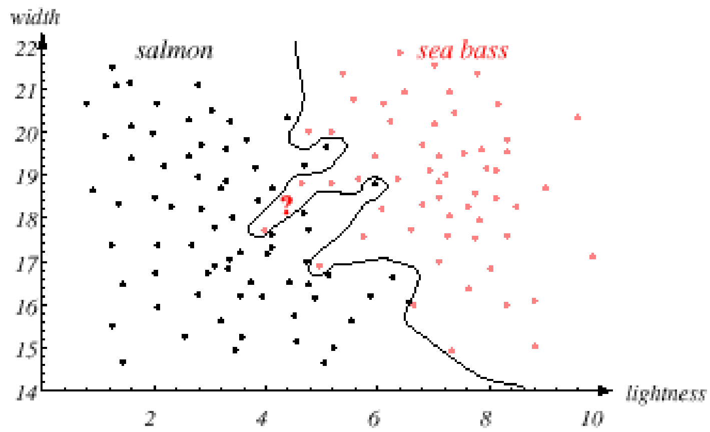
Two Features of Lightness and Width



Generalization and Decision Boundary

- We might add other features that are not correlated with the ones we already have, e.g. shape parameters.
- A precaution should be taken not to reduce the performance by adding redundant features.
- Ideally, the best decision boundary should be the one which provides an optimal performance such as in the following figure:

Complex Decision Boundary



Generalization and Decision Boundary

- However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify novel input

Generalization!

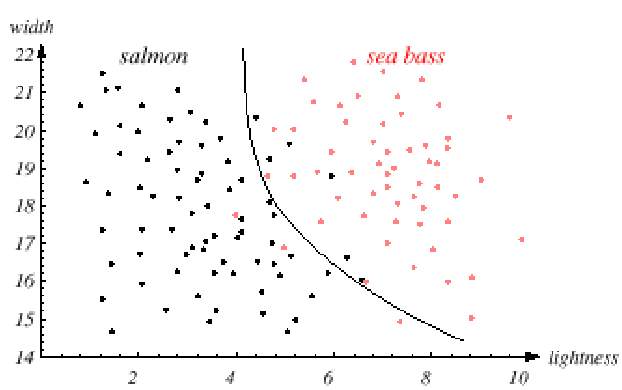
- It is unlikely that the complex decision boundary would provide good generalization.

Generalization and Decision Boundary

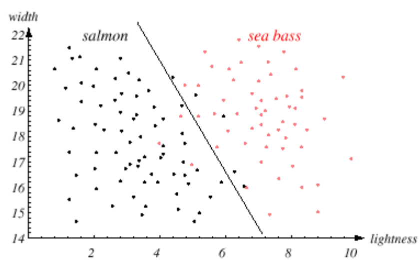
- More training samples for estimating the true characteristics of the categories.
- Amount of data in most problems is limited.
- Even with vast data, the classifier can give a complicated decision boundary.
- A simple classifier with non-complex decision boundary can provide good generalization.

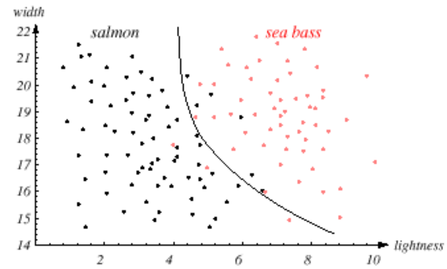
Selected decision boundary

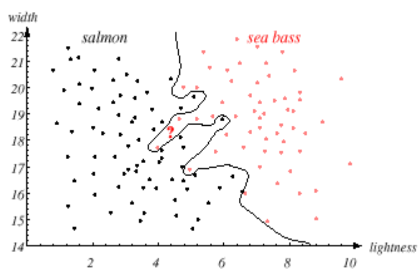
- The decision boundary can be a simple curve which might represent the optimal trade-off.



Decision Boundary Selection





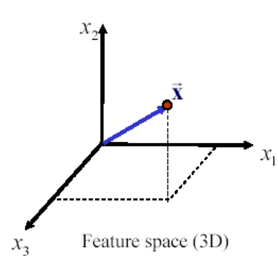


- Which of the boundaries would you choose?
  - *Simple linear boundary – training error > 0*
  - *Nonlinear complex boundary – training error = 0*
  - *Simpler nonlinear boundary – training error > 0*

Terminologies in Machine Learning

- Features: a set of variables believed to carry discriminating and characterizing information about the objects under consideration
- Feature vector: A collection of *d* features, ordered in some meaningful way into a *d-dimensional* column vector, that represents the signature of the object to be identified.
- Feature space: The *d-dimensional* space in which the feature vectors lie. A *d-dimensional* vector in a d-dimensional space constitutes a point in that space.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \begin{array}{l} \textit{feature 1} \\ \textit{feature 2} \\ \\ \textit{feature d} \end{array}$$



Feature space (3D)

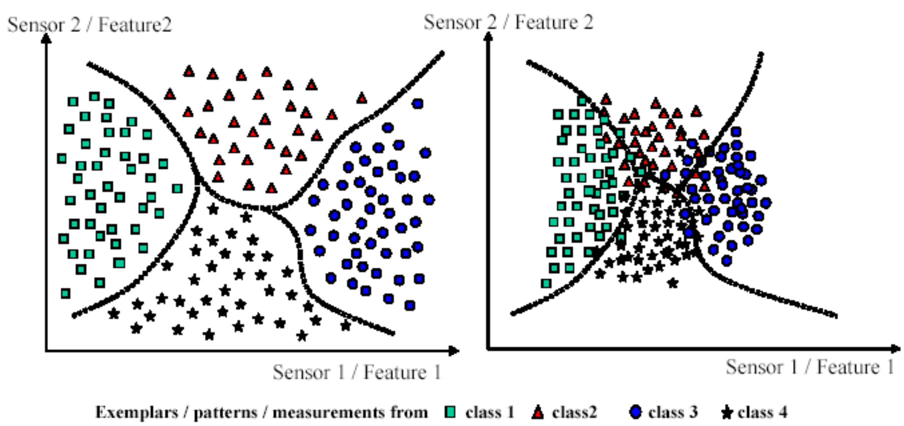
Terminologies in ML

- Class: The category to which a given object belongs
- Decision boundary: A boundary in the *d-dimensional* feature space that separates patterns of different classes from each other
- Training Data: Data used during training of a classifier for which the correct labels are *a priori* known
- Testing Data: Unknown data to be classified. The correct class of this data are not known *a priori*

Terminologies in ML

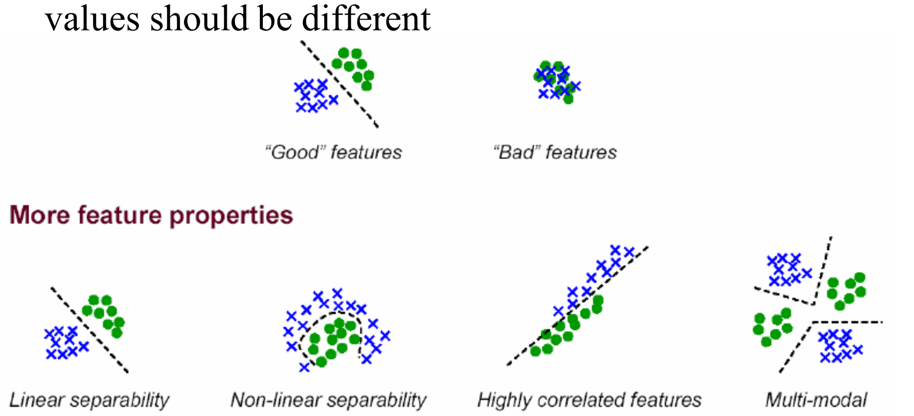
- Classifier: An algorithm which adjusts its parameters to find the correct decision boundaries –through a learning algorithm using a training dataset
- Error: Incorrect labelling of the data by the classifier
- Training Performance: The ability/performance of the classifier in correctly identifying the classes of the training data, which it has already seen. It may not be a good indicator of the generalization performance.
- Generalization (Test Performance): The ability/performance of the classifier in identifying the classes of previously unseen

Kinds of Data



Good Features vs. Bad Features

- Ideally, for a given group of patterns coming from the same class, feature values should all be similar
- For patterns coming from different classes, the feature values should be different



Components of Machine Learning System

What is the cost of making a decision –what is the risk of an incorrect decision?

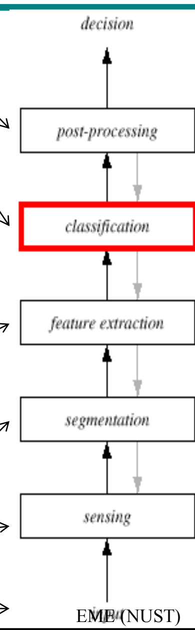
Determine the right type of classifier, right type of training algorithm and right type of classifier parameters. How to choose a classifier that minimizes the error and maximizes the performance?

Identify and extract features relevant to distinguishing sea bass from salmon, but invariant to noise, occlusion, rotations, scale, shift. What if we already have irrelevant features? Dimensionality reduction?

Separate & isolate fish from background image

Camera on a conveyor belt

Fish



Machine Learning Systems

- Sensing
  - *Use of a transducer (camera or microphone).*
  - *ML system depends on the bandwidth, resolution, sensitivity, distortion, etc. of the transducer.*
- Segmentation
  - *Patterns should be well separated and should not overlap.*
- Feature extraction
  - *Distinguishing features*
  - *Invariant features with respect to translation, rotation and scale.*

Machine Learning Systems

- Classification
  - *Use a feature vector provided by a feature extractor to assign the object to a category.*
  - *Not always possible to determine the values of all the features.*
- Post Processing
  - *Post-processor uses the output of the classifier to decide on the recommended action.*
  - *Error rate*

Learning and Adaptation

- Learning incorporates information from training samples in classifier design.
- It refers to some form of algorithm for reducing the error on training data.

- Supervised learning
  - *A teacher provides a category label for each pattern in the training set.*

Learning and Adaptation

- Unsupervised learning
  - *The system forms clusters or "natural groupings" of the input patterns.*
  - *The labels of the categories are unknown.*

- Reinforcement Learning
  - *Learning with a critic.*
  - *No desired category signal is given; instead the only teaching feedback is that the tentative category is right or wrong.*

Conclusion

- Overwhelmed by the number, complexity and magnitude of the sub-problems of Machine Learning.

- Many of these sub-problems can indeed be solved.

- Mathematical theories solving some of these problems have in fact been discovered.

- Many fascinating unsolved problems still remain.