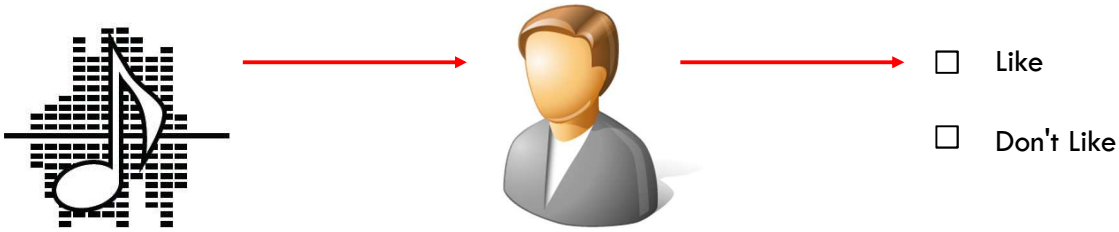
MACHINE LEARNING FOR DATA MINING LECTURE 2-1: SCATTER PLOTS, DECISION SURFACES

Siamak Sarmady (Urmia University of Technology)
Sebastian Thrun, Katie Malone (Google)

Inductive (Supervised) Learning

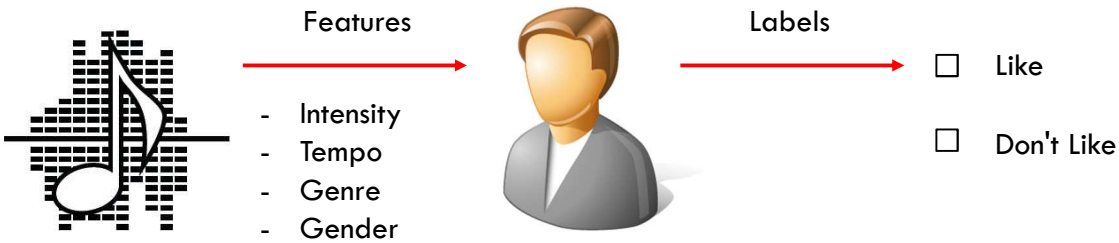
- Given examples of a function (X, F(X))
- Build a function (or model) that can predict F(X) for new examples X
  - **Discrete F(X) :** Classification
  - **Continuous F(X) :** Regression
  - **F(X) = Probability(X) :** Probability estimation
    - in fact a regression with output value in the range of [0,1]

Music Suggestion



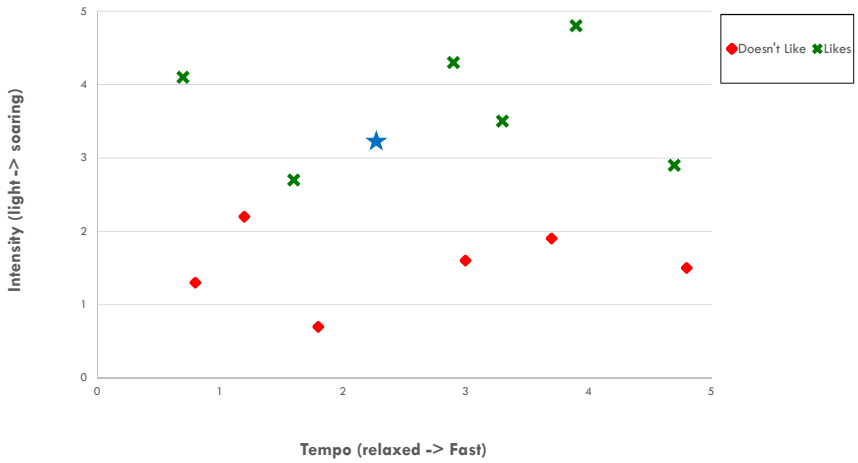
Assume we have a list of 100 music labels someone has listened and liked or disliked (the person specifies that for every label). Now we want to predict whether the person will like or don't like a new music item.

Music Suggestion



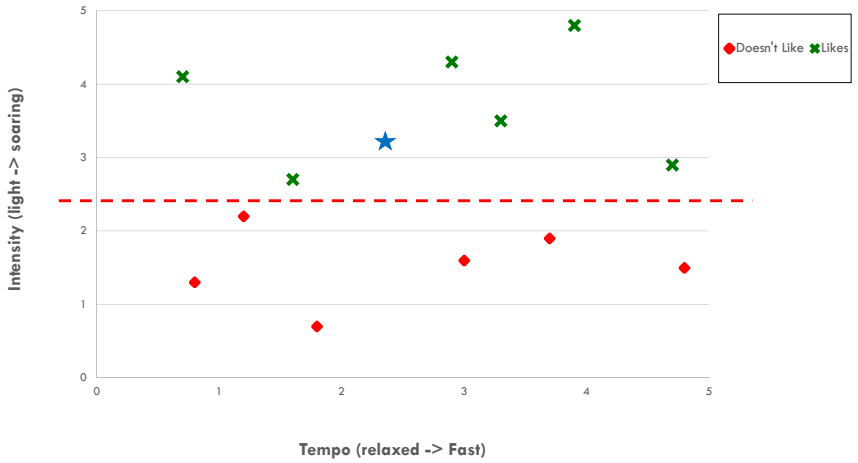
We may select a few effective parameters. We then determine those parameters for the 100 items the person liked/disliked. Using those parameters we train a classifier and try to guess whether a new music is liked or disliked by the person.

Scatterplot for 2 features (Kate)



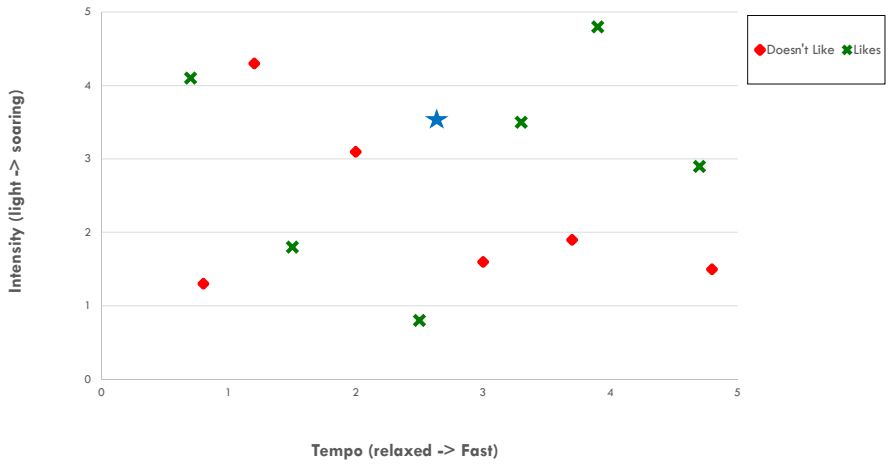
- Assume we have a candidate music specified with the star. Will Kate like that music?

Scatterplot for 2 features (Kate)



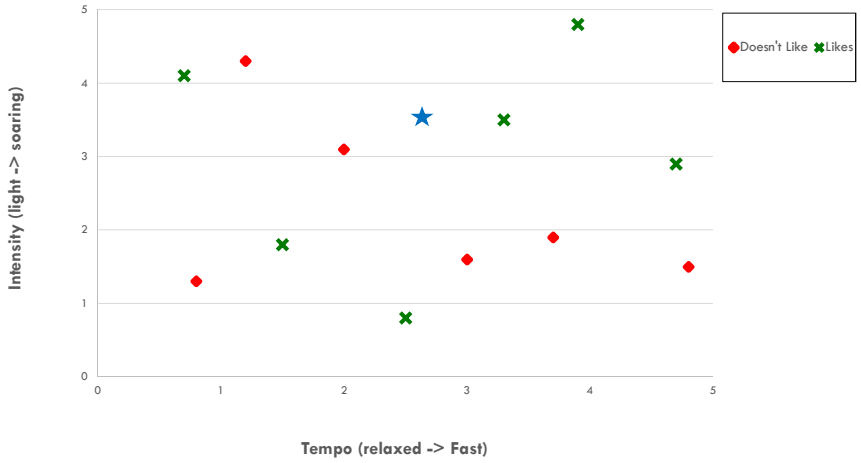
- It seems Kate likes high intensity songs (weather relaxed or fast tempo)
- We cannot be sure but we can estimate the boundry of the two classes using that line.

Scatterplot for 2 features (John)



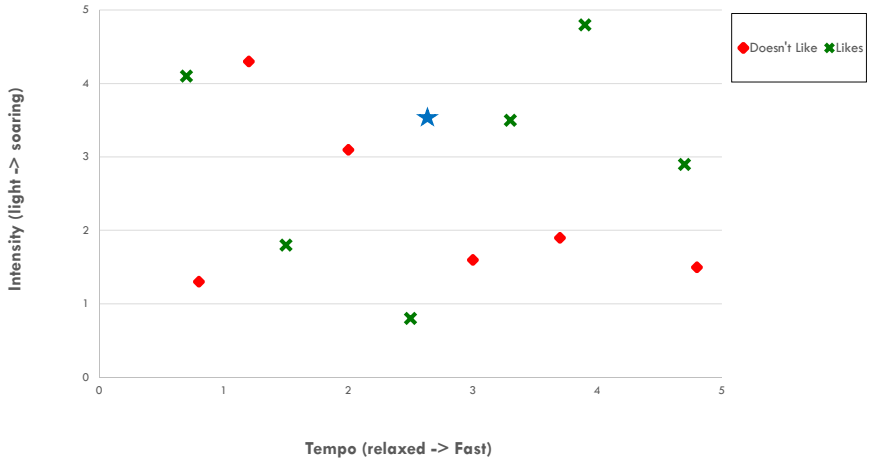
- This time, John's interest is more scattered. We cannot easily specify a boundry.
- We cannot be sure but we can estimate the boundry of the two classes using that line.

Scatterplot for 2 features (John)



- The star might be in the green area

Scatterplot for 2 features (John)



- Or in the red area. It is not clear!
- Classification or regression

Autodriving Car – Speed Adjustment / Limiting

Bumpiness

Smooth

Bumpy

Very Bumpy

Slope

Flat

Slope

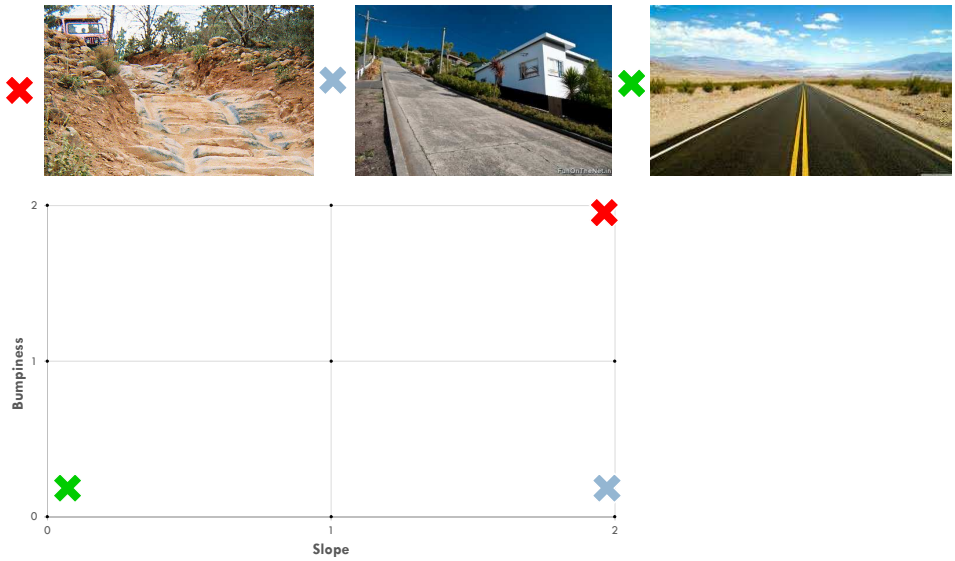
Very Steep

- Assume we want to consider two parameters (bumpiness and slope) for speed limit adjustments

Scatter Plot

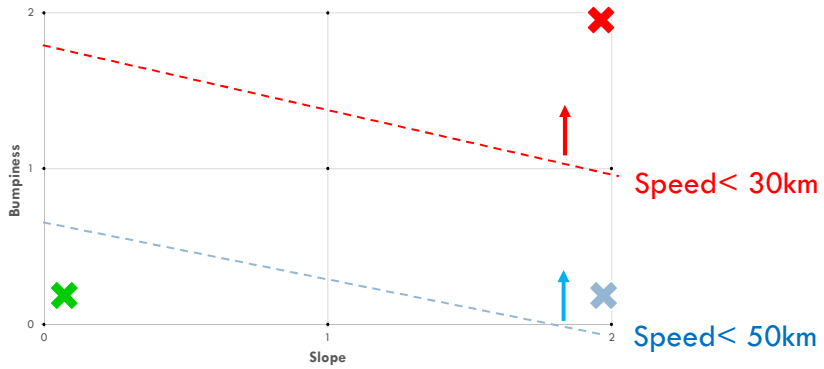
- It is possible to have a mix of the three slope and bumpiness levels

Scatter Plot



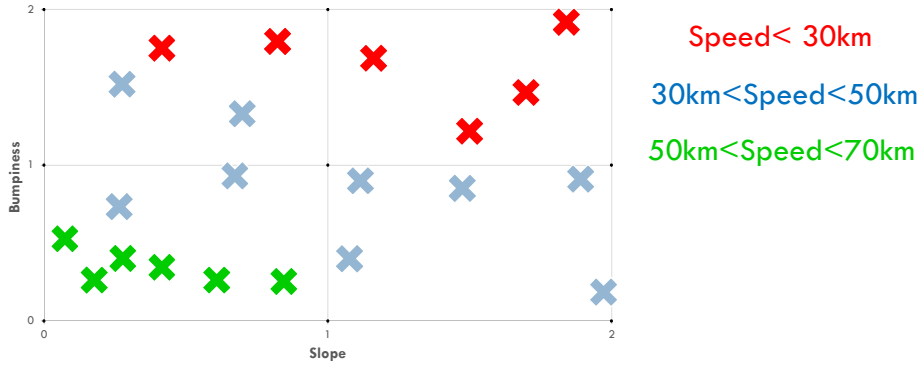
- The place of each case is determined by both bumpiness and slope

Speed Limiting – without learning



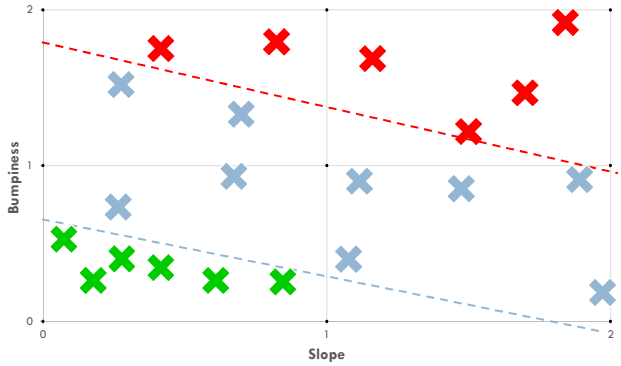
- Based on the area of the scatter plot we may determine the proper speeds. We call these lines 'Class Boundries' or 'Decision surface'. We can also design a discrete speed limit function like v = f(S,B) ...
- In this example, we manually design the behaviour of the car...

Speed Limiting – with learning



- If we want to use learning, we drive the car in different places and situations, with proper speeds. The sensors gather slope, bumpiness and speed levels (i.e. One of 3). We then use the data for learning.
- **Note:** we have decided to have 3 different speed limits even before performing the experiments

Speed Limiting – with learning



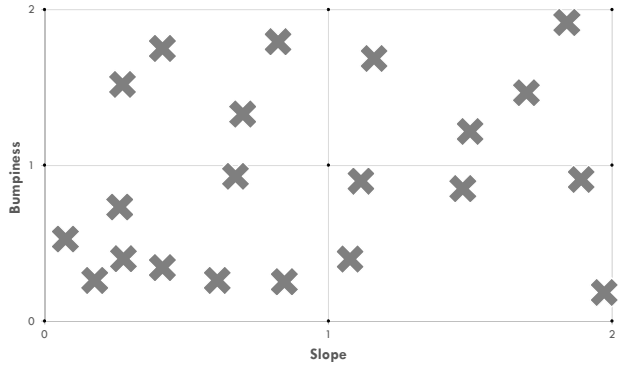
Speed< 30km

30km<Speed<50km

50km<Speed<70km

- What machine learning is doing, is to find decision surfaces that separates classes
- Based on those surfaces (and the sensor values for slope and bumpiness), the car decides on a proper speed limit

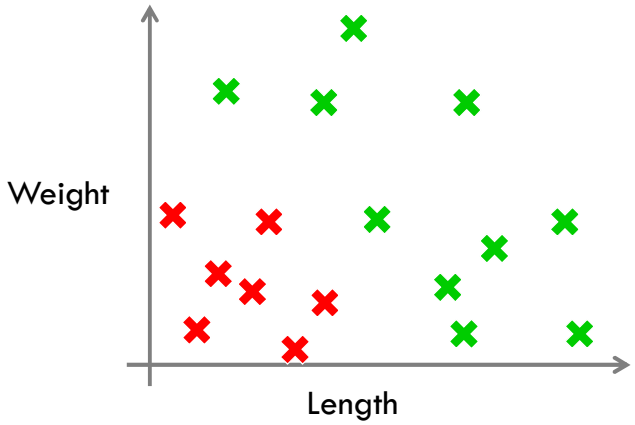
Cruise Speed Adjustment



- What if we wanted the mechanism to determine the proper cruise speed (Instead of speed limit ranges)?
- We would need the speed of each of those data points (i.e. continuous) instead of just speed level (discrete class).
- We then would need the supervised learning algorithm to solve a regression problem and estimate a function for us i.e. v = f(s,b). The function would then calculate the proper speed for every continuous value of slope and bumpiness
- We would need a 3rd dimension to show the speed function.

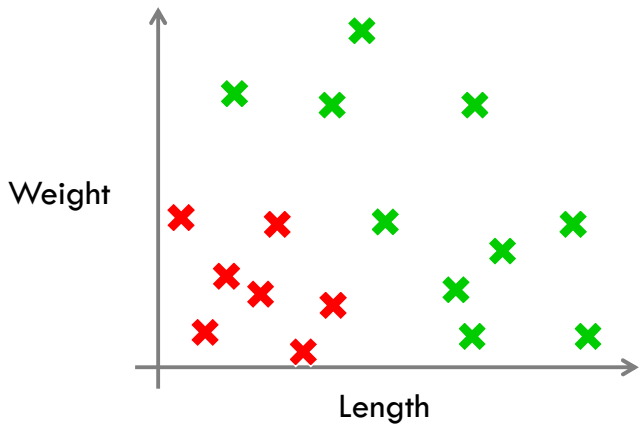
Decision Surface

- What machine learning does, is to define a decision surface which separates the data into two or more classes.
- Assume we have 1000 kg of Tuna fish. We want to select quality ones and send them to market. A worker selects fishes manually and we record the parameters and his decisions.



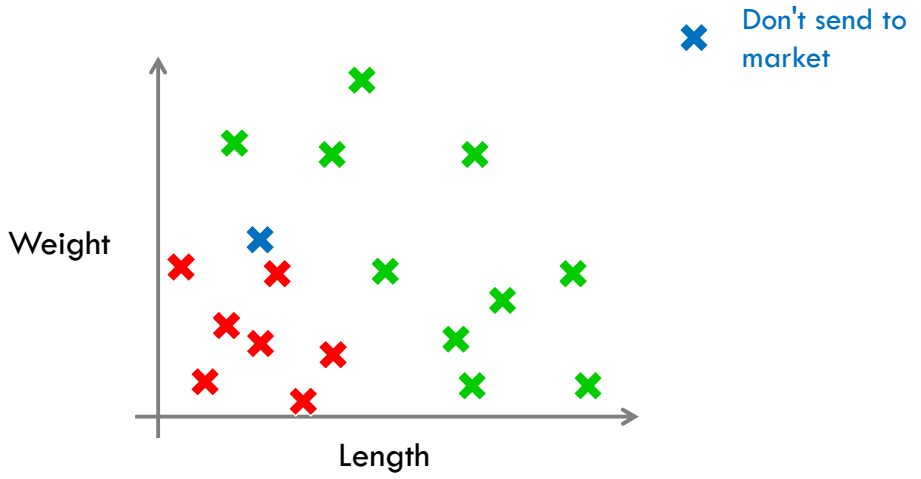
Decision Surface

- Based on learning data, the machine will select those fishes that fall in green section. Those in red section are too small to be sold.



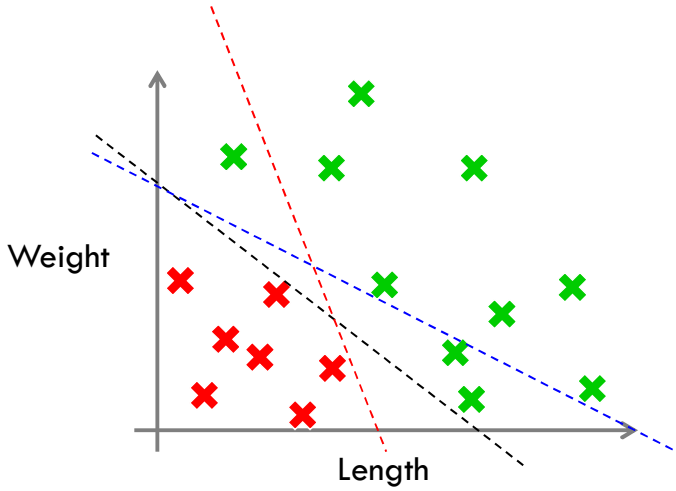
Decision Surface

- Now the decision boundry can be used for the decision making about the future fishes.
- We actually generalize what we learned to unseen cases.



Linear Decision Surface

- If the decision surface is a line, we call it a linear decision surface. You may calculate different lines for the purpose. Which one is the best? What is a good decision surface?



- Red line misclassifies 1 point, it is to near to red data points
- Blue line missclassifies multiple data points
- The **black line** seems more ok

Good Decision Surface

- A good decision surface is the line which gives low error for the learning data

  as well as

- it generalizes the learned knowledge in a way that the average error is lowest for the unseen data.

- **Classification:**
  - Data -> Decision Surface -> Predictions (Generalization)

- **Regression:**
  - Data -> Function (model) -> Predictions (Generalization)