Reading 7

☆ Video: SVMs (Support Vector Machines)

- Imagine you measure mass of many mice.
- You label each mice as obese or not obese.

- Focus on the observations of the two different groups that are closest to each other and have the midpoint be the treshold.

- The shortest distance between the observations and the threshold is called the margin.

- When we use the threshold that gives us the largest margin to make classifications, we are using a maximal margin classifier.
  - ↳ These are really sensitive to outliers since it does not allow for misclassifications.

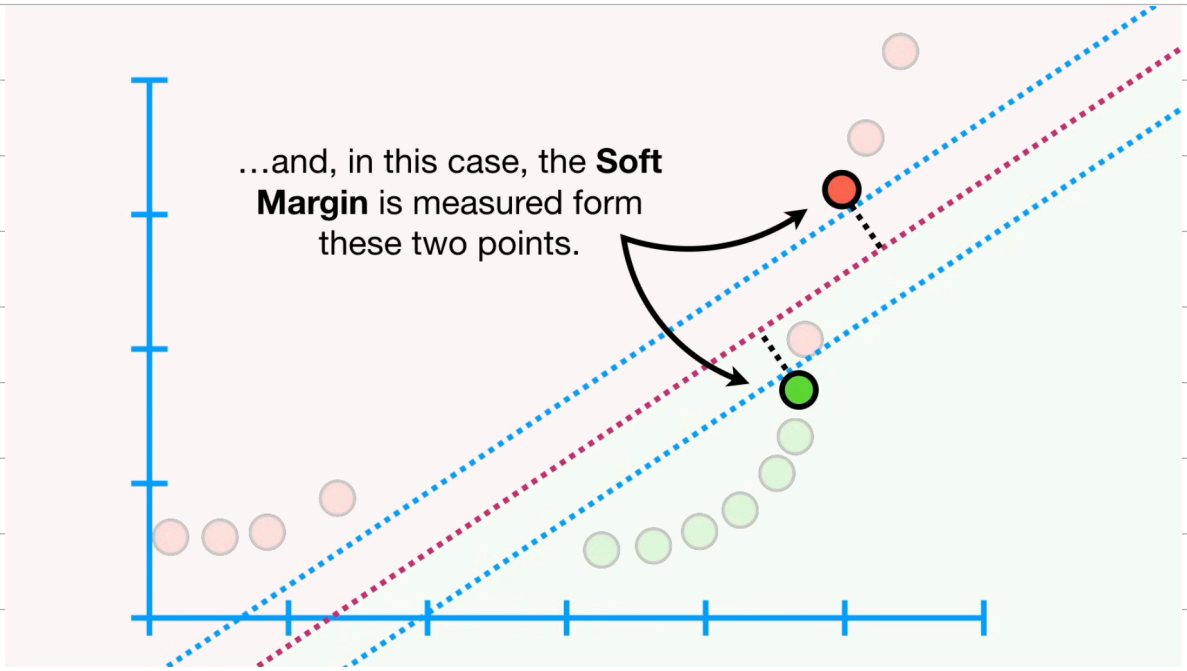
- When we allow misclassifications, the distance between the observations (of different classes) and the threshold is called a Soft Margin.

- We use cross validation to determine how many misclassifications and observations to allow inside of the Soft Margin to get the best classification.

- When we use a soft margin to determine threshold we are using a soft margin classifier a.k.a. Support vector classifier. (SVC)

- The observations on the edge and within the soft margin are called Support vectors.

- If data is 2D, the support vector classifier is a line.

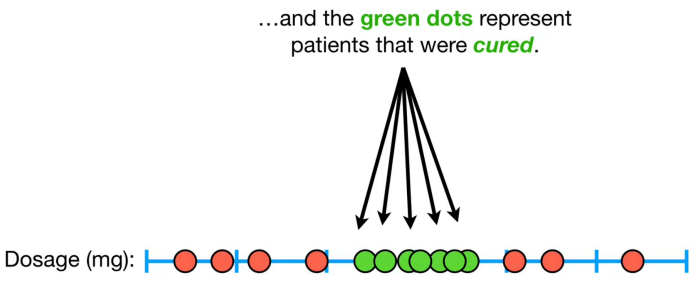


- If the data is 3D, the SVC forms a plane.

- If the the data is 4D or more the SVC is a hyperplane.

- SVCs allow for outliers and misclassifications.

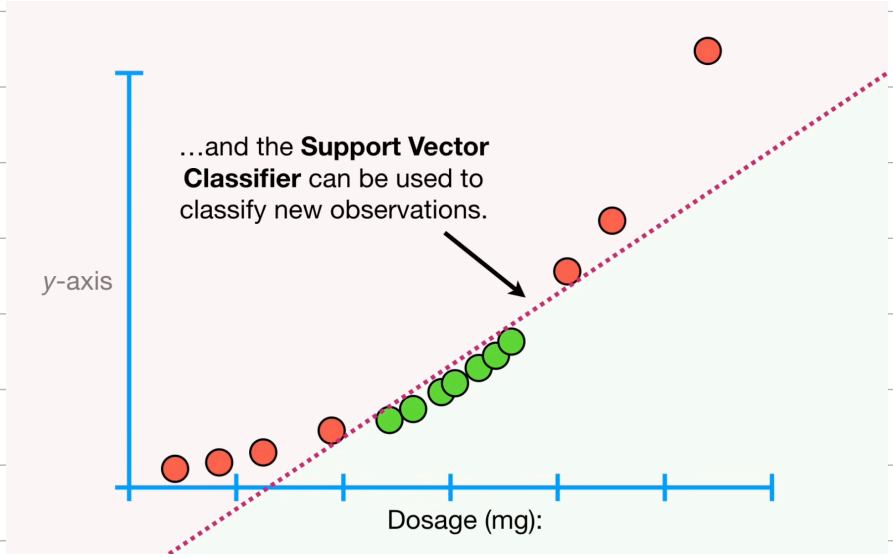
- But what if we had data like this:



- SVCs don't work here.

- We need to use Support Vector Machines (SVMs)

- The SVM would add a Y-axis that is $(dosage)^2$
  - ↳ Now you can draw a SVC to seperate cured vs not cured:



- The main ideas of SVMs:
  - ↳ Start with data in low dimension
  - ↳ move the data to a higher dimension
  - ↳ find a SVC that splits the data into two groups.

- How do we decide how to transform the data?

  - Kernel functions systematically finds SVCs that work in higher dimensions.

    - ↳ A polynomial kernel systematically increases dimensions by setting hyperparameter d, the degree of the polynomial and the relationships between each pair of observations are used to find a SVC.

  - ↳ We can find a good value for d with cross validation.

- Another common kernel is the Radial Kernel, also known as the Radial Basis Function (RBF) Kernel.
  - ↳ It finds SVCs in infinite dimensions
  - ↳ It behaves like weighted nearest neighbor model.
    - ↳ so observations closer to new observation have more weight on the classification of the new observation.

- Kernel functions only calculate the relationships between every pair of points as if they are in the higher dimensions; they don't actually do the transformation.
  - ↳ This is known as the kernel trick.
  - ↳ It reduces the amount of computation needed for SVMs since no transformations need to be computed.
  - ↳ It makes calculating relationships in infinite dimensions used by the radial Kernel (RBF) possible.