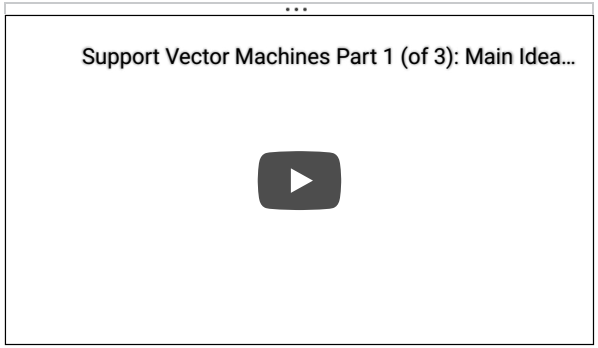
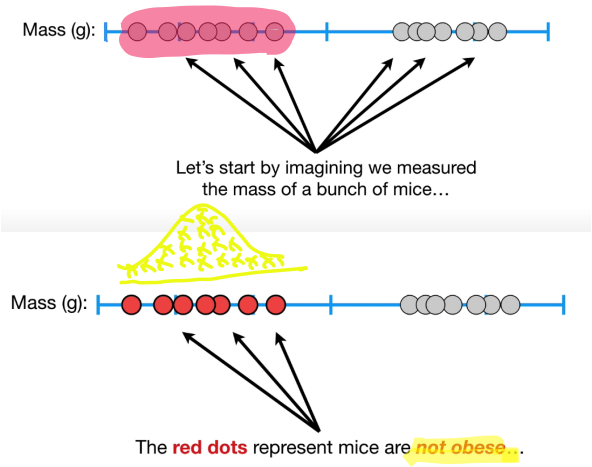
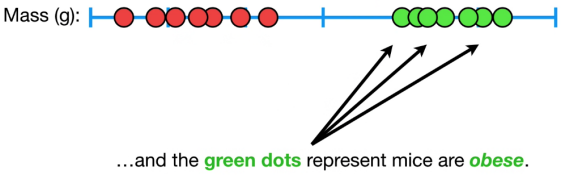
5 - Support Vector Machine - Overview

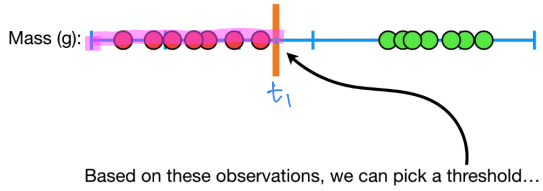
Sunday, 22 November BE 2563 19:41

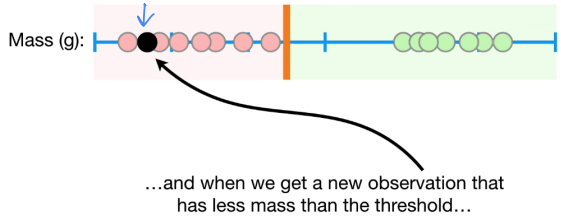
External Resources: Support Vector Machines, Clearly Explained!!!

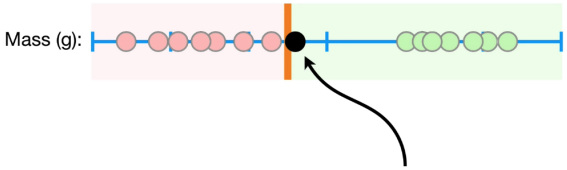




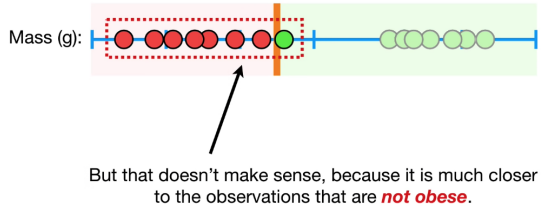


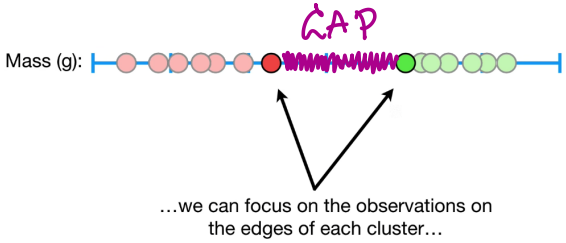


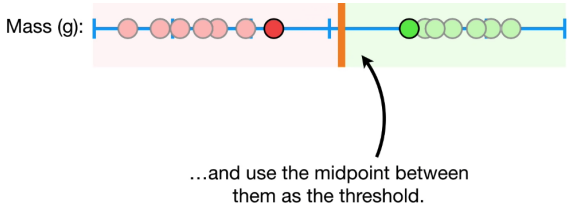


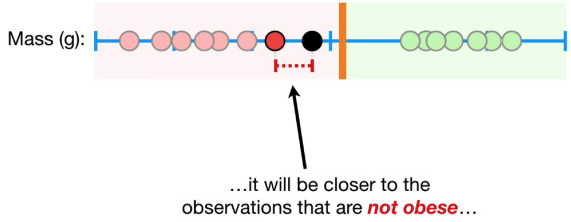


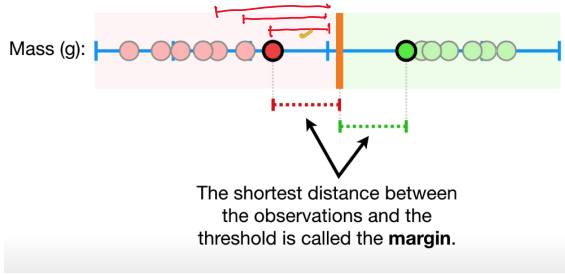
However, what if get a new observation here?

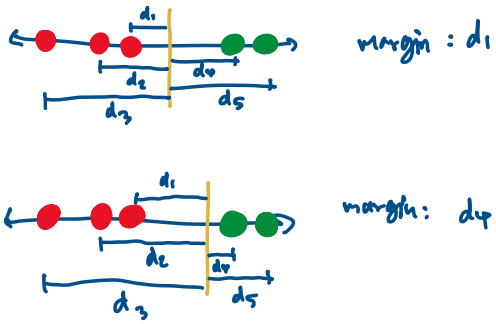


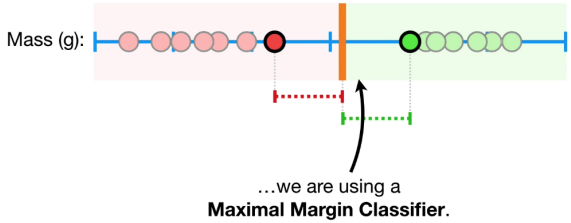


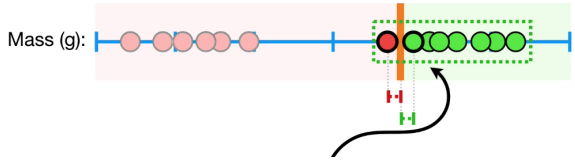

Mass (g):
…and use the midpoint between
them as the threshold.


Mass (g):
…it will be closer to the
observations that are *not obese*…

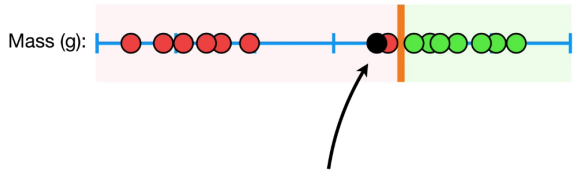

Mass (g):
The shortest distance between
the observations and the
threshold is called the **margin**.


margin : d_1
margin : d_4


Mass (g):
…we are using a
Maximal Margin Classifier.

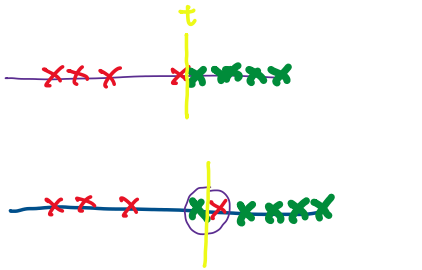

Mass (g):

In this case, the **Maximum Margin Classifier** would be super close to the ***obese*** observations...



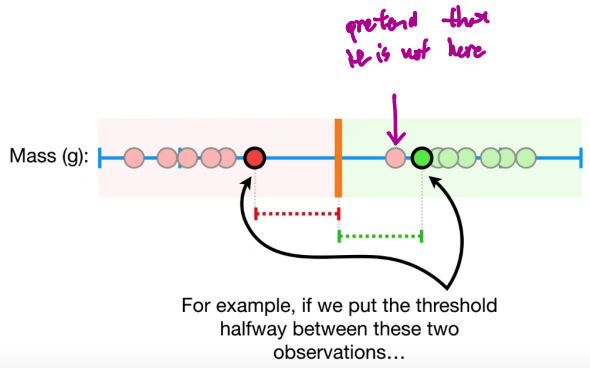
Now, if we got this new observation...

...we would classify it as ***not obese***, even though most of the ***not obese*** observations are much further away than the ***obese*** observations.

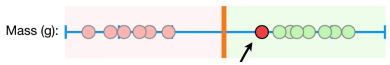


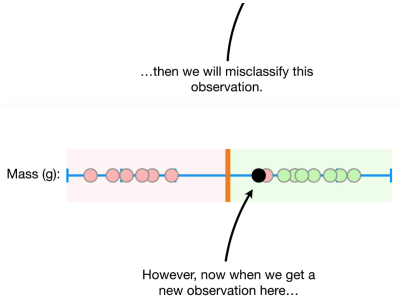
So **Maximal Margin Classifiers** are *super sensitive to outliers* in the training data and that makes them pretty lame.

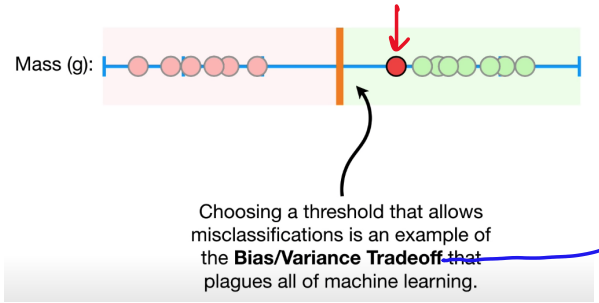
To make a threshold that is not so sensitive to outliers we must **allow misclassifications**.



For example, if we put the threshold halfway between these two observations...







BIAS vs VARIANCE

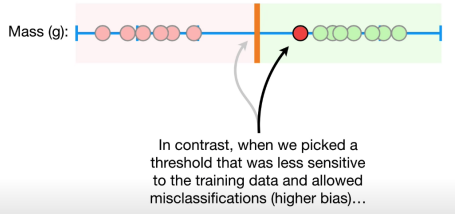
low bias = we just have to trust all the data points to do the job

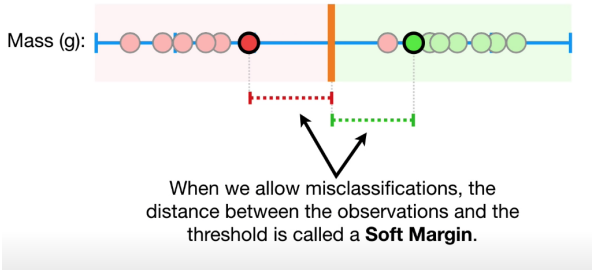
high bias = we have some concerns/suggestions about the data points

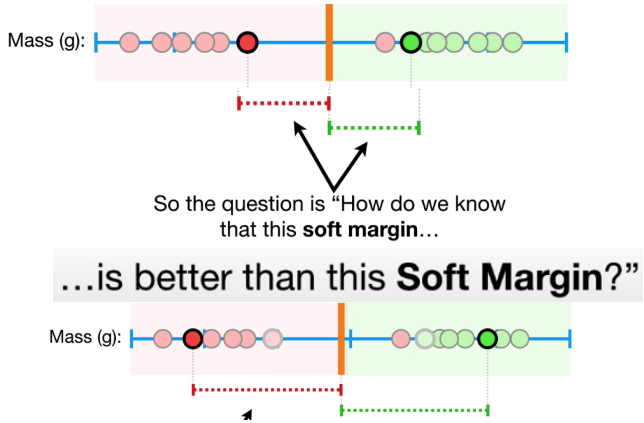
In other words, before we allowed misclassifications, we picked a threshold that was very sensitive to the training data (low bias)…

B↓ V↑

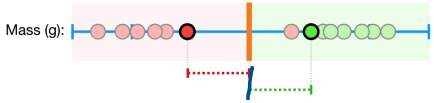
B↑ V↓

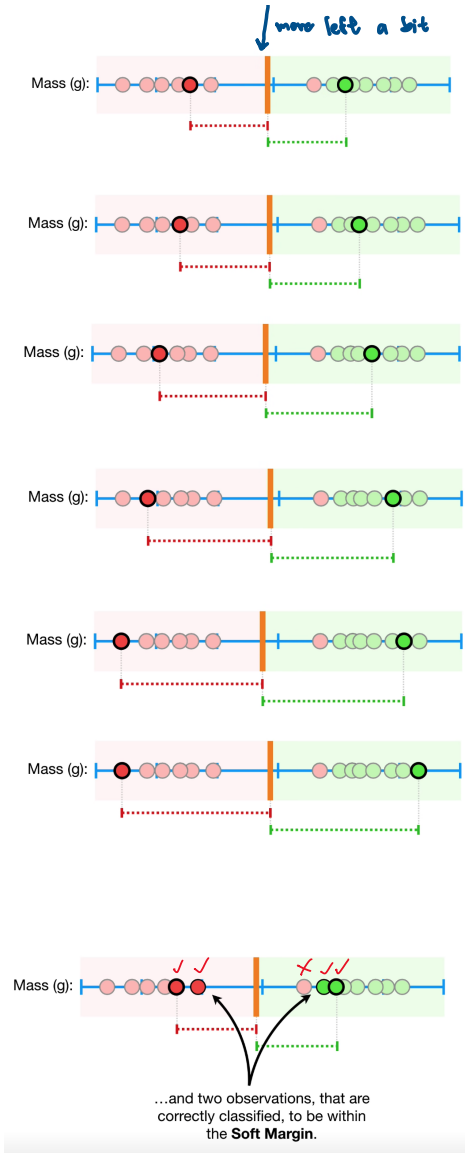


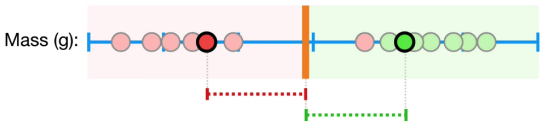




The answer is simple: We use **Cross Validation** to determine how many misclassifications and observations to allow inside of the **Soft Margin** to get the best classification.





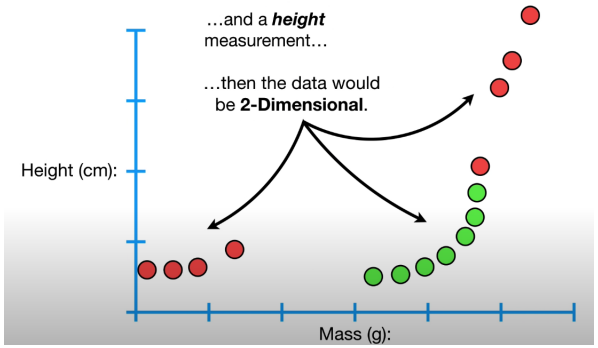


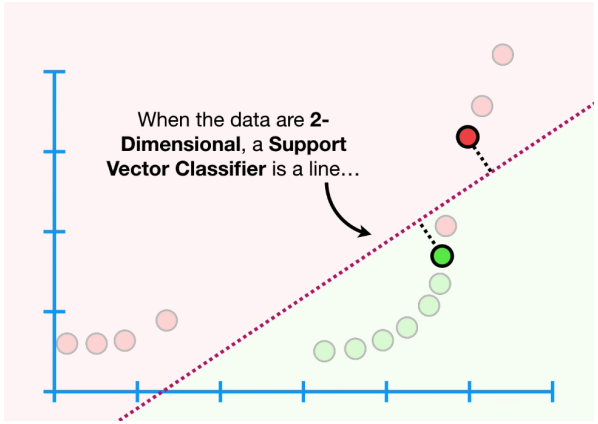
When we use a **Soft Margin** to determine the

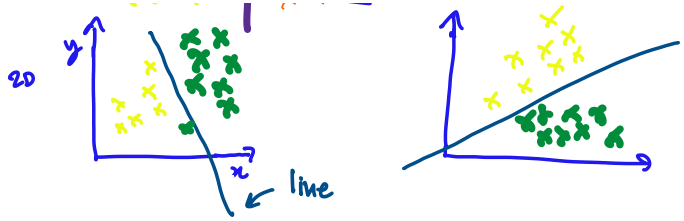
location of a threshold…

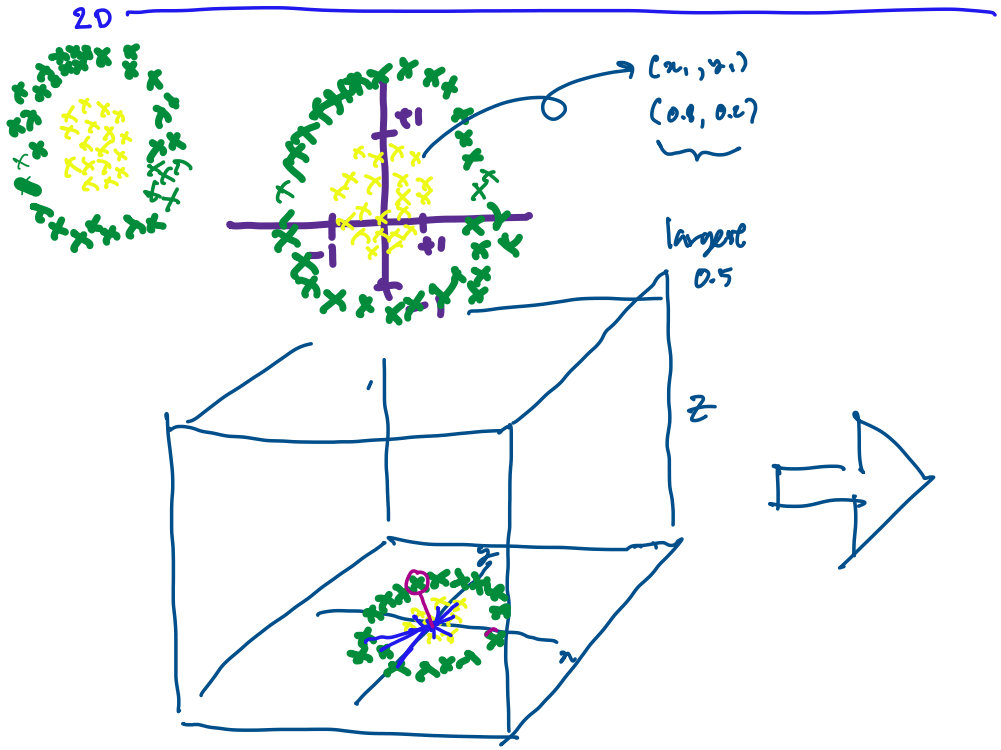
…then we are using a **Soft Margin Classifier** aka a **Support Vector Classifier** to classify observations.

The name **Support Vector Classifier** comes from the fact that the observations on the edge *and within* the **Soft Margin** are called **Support Vectors**.

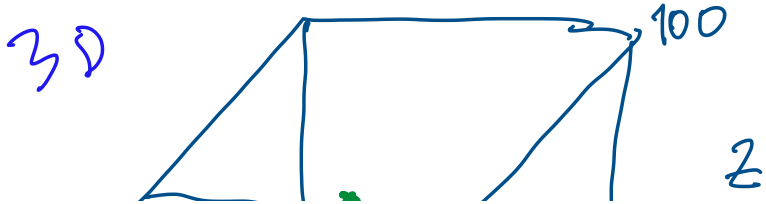


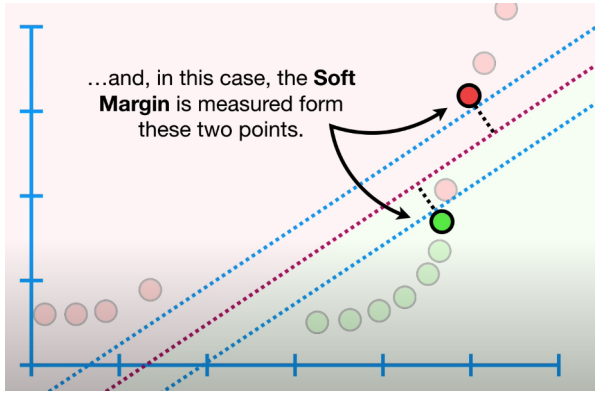


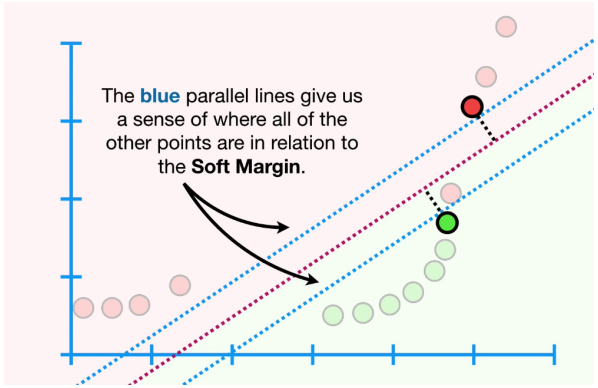


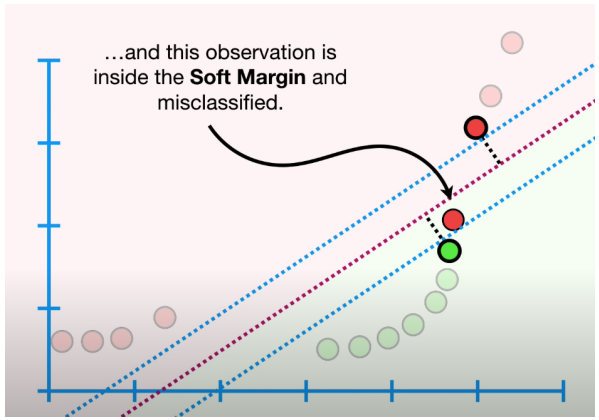


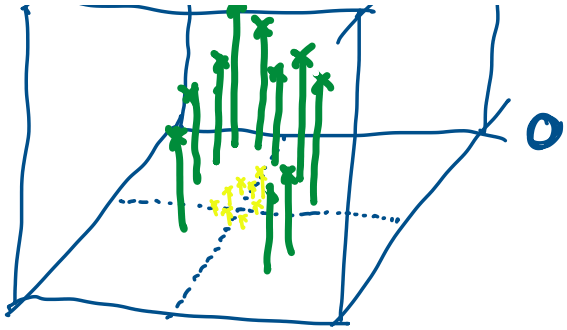
$z = d^2$

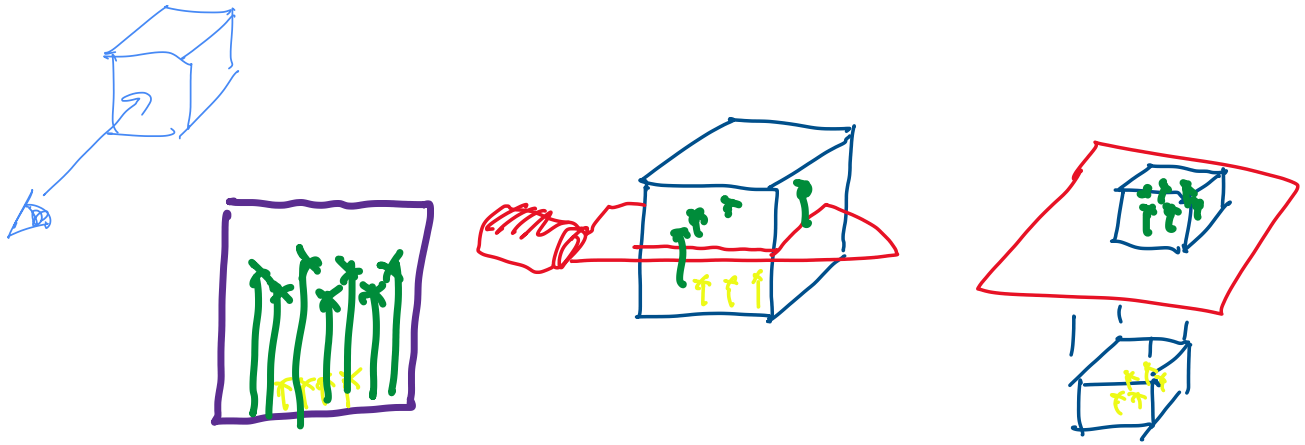



…and, in this case, the **Soft Margin** is measured form these two points.

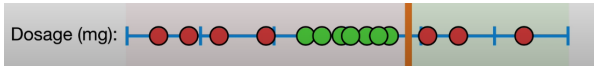

The **blue** parallel lines give us a sense of where all of the other points are in relation to the **Soft Margin**.


…and this observation is inside the **Soft Margin** and misclassified.



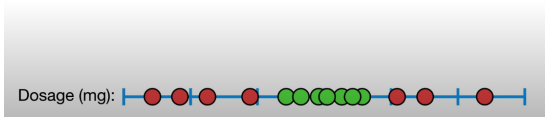


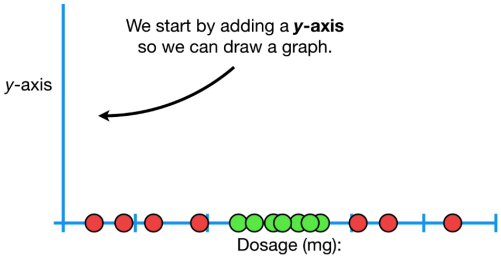
SVC.... Has serious limitation

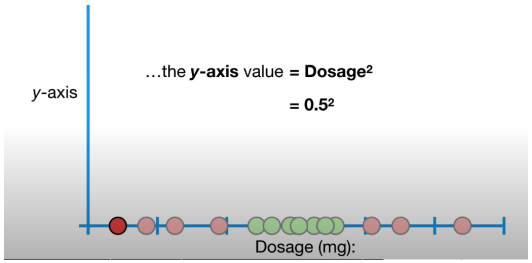


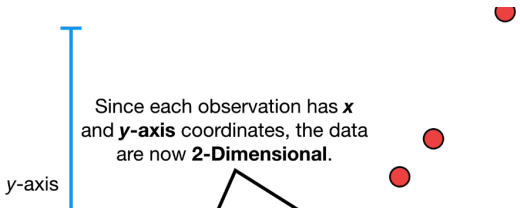
Since **Maximal Margin Classifiers** and **Support Vector Classifiers** can't handle this data, it's high time we talked about…

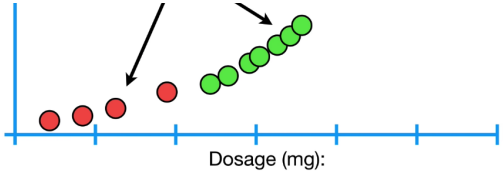
Support Vector Machines!!!

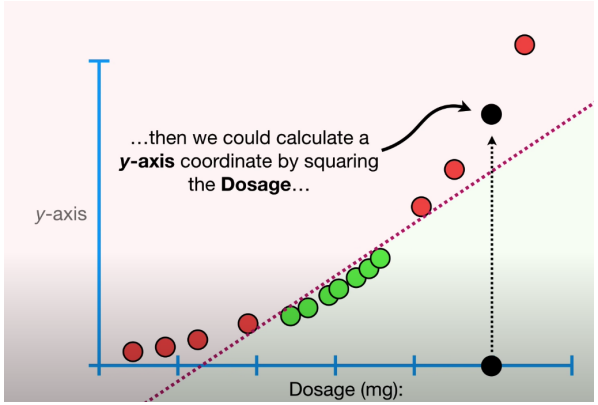


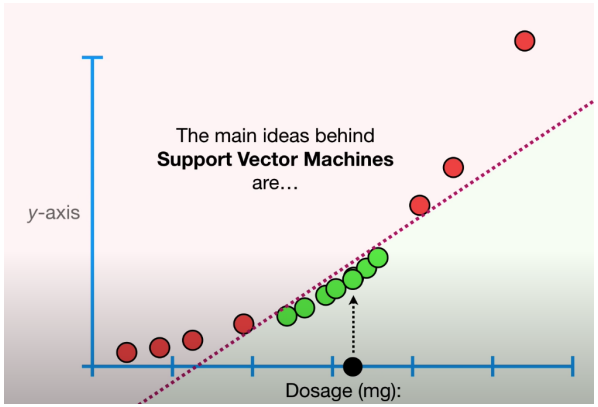


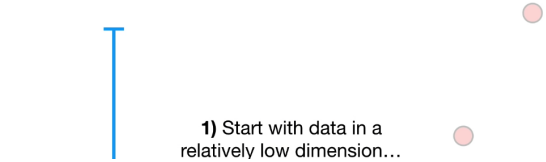


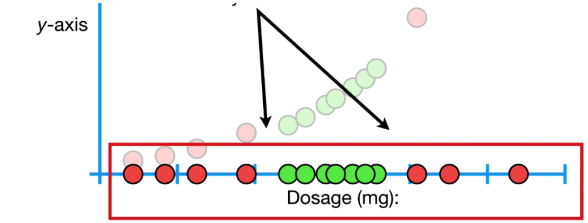


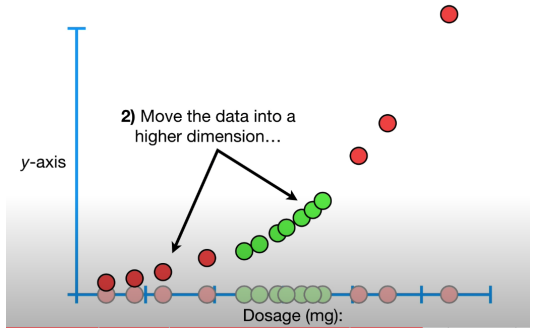

Dosage (mg):

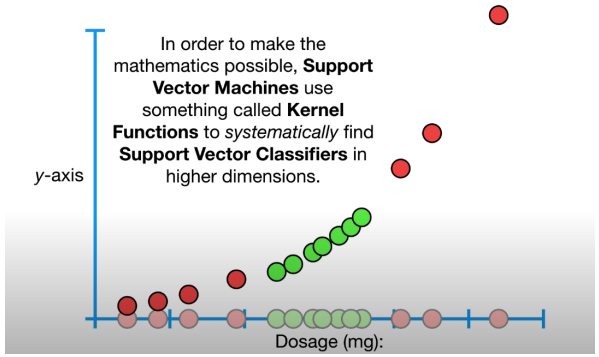

…then we could calculate a
y-axis coordinate by squaring
the **Dosage**…
y-axis
Dosage (mg):


The main ideas behind
Support Vector Machines
are…
y-axis
Dosage (mg):


1) Start with data in a
relatively low dimension…

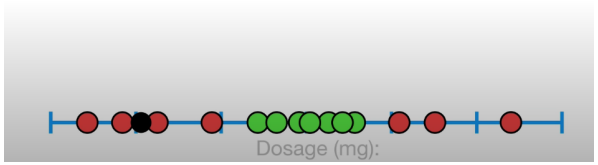






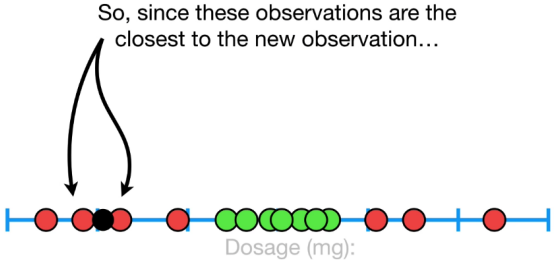
Another very commonly used **Kernel** is the **Radial Kernel**, also known as the **Radial Basis Function (RBF) Kernel**.

…the **Radial Kernel** behaves like a **Weighted Nearest Neighbor** model.



In other words, the closest observations (aka the nearest neighbors) have a lot of influence on how we classify the new observation…

So, since these observations are the closest to the new observation…



…the **Radial Kernel** uses their classification for the new observation.