Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

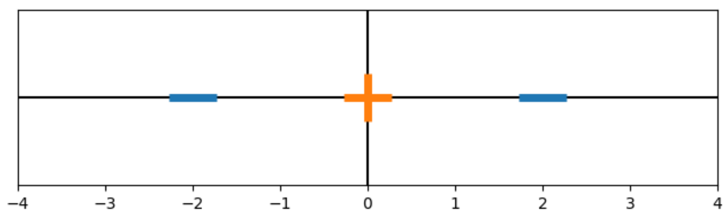
Fall 2022

Gaining intuition about feature transformations

Now that we understand the general idea of feature transformations, we can begin work with them, particularly for classification.

Our goal is often to take data that linear models couldn't handle, and make it **separable.**

So, we'll consider maybe the simplest (solvable) case of a nonlinear data set:



In its current state, there's no one plane that would go through these data points: this is where our transform comes in. We'll try using a polynomial transform with x^2. It turns out, $-x^2 + 2$ works pretty well.

How do we visualize this? It turns out, there are different perspectives:

> **Clarification 1**
>
> There are **two** different ways we can **graph** a transformation:
>
> - We transform the **separator**: if our model is $-x^2 + 2$, we just graph that function over the data.
>   - This is the approach we used above: we wanted a line that **fit** to our data.
> - We transform the **data**: we graph each data point according to our model $-x^2 + 2$.
>   - This model allows us to keep a "**linear** separator": we effectively "shift" the data to be linear.
>
> These models are mathematically **equivalent**, and we'll switch the approach we're using based on which is easier/more useful to graph.

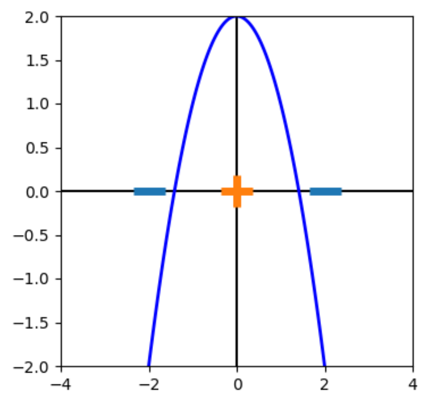
It may seem concerning to transform the data, rather than the model. However, keep in mind that:

- If you switch perspectives, they're the same: a different model will just "transform" the data when calculating its output.
- Usually, we try to preserve the original structure of the data, so we don't lose information: we just add more.
  - For example, $[1, x, x^2]$ still contains the information x: we just add x^2.

Example: Let's show both of these in action.

Transforming our separator

First, we transform our linear separator as desired: graphing $-x^2 + 2 = 0$ on our plot.



In this version, we've taken our hyperplane separator and transformed it nonlinearly.

In this case, we have assigned $-x^2 + 2 < 0$ as positive.

In this version, we preserve the structure of the data, making it easier to see the original shape. However, it's not as easy to think about the shape and orientation of the "plane" now that it's been deformed.

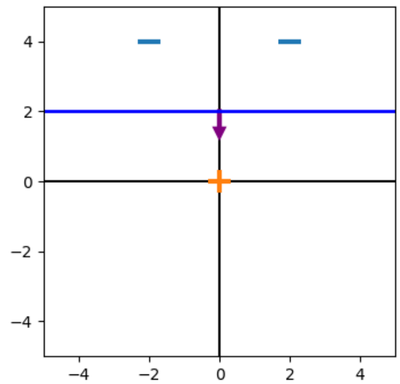
For example, we don't really have a good "normal" vector, even if we know which side is positive. This is why, to keep our model "linear", we can transform the data, and find the corresponding plane. We'll do that next.

Transforming our data

In this case, every data point gets plotted on $[x, x^2]$. Our hyperplane is given by

$$-x^2 + 2 = \overbrace{\begin{bmatrix} 0 \\ -1 \end{bmatrix}^T}^{\theta^T} \begin{bmatrix} x \\ x^2 \end{bmatrix} + 2 \tag{1}$$

Thus, we get a θ pointing downward, with an offset of 2.



This time, we've transformed our data: the math is totally the same, but now we can identify our separator more easily.

Note that our transformation makes the data linearly separable!

Concept 2

Features transformations allow us to **non-linearly** transform our data, in order to make that data **linearly separable**, or at least, more **accurate** with a linear separator.

Often, we do this by transforming into a **higher dimensional** space.

Positive vs. Negative

While these perspectives are helpful, they can become too complicated with more dimensions/higher-dimensional transformations.

In an effort to simplify, we might ask ourselves, "what do we really want to know"? In the end, all we typically care about is classification: which data points are positive or negative?

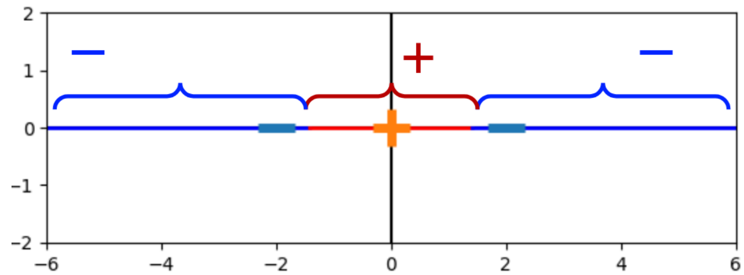
So, we'll create a third representation to correspond to that.

Concept 3

A third, **simplified** representation of our transformation doesn't show how it affects our data points or classifier. Instead, we just show the **result**: which regions are classified as positive, and which are classified as negative?

This allows us to see which points get **classified** in which way, without considering the high-dimensional details of the model itself.

Example: We can graph this for our sample data:



This way, we can stay in a 1-D space, while showing the information we need!

Note that the points where we switch between positive and negative, $\pm\sqrt{2}$, are the points corresponding to $-x^2 + 2 = 0$: they're the only part of the separator surface visible in our 1D plot.

They match our nonlinear hyperplane separator from section* 5.1.1