Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

Fall 2022

Radial Basis

Finally, we consider an alternative way to create a feature space.

- With the "polynomial basis" approach, we **combined features** to create more complex surfaces to **fit** the structure of the data.
- This "radial basis" approach, on the other hand, **combines data points** to **learn** more about the structure of the data.

What do we mean by that? Well, let's consider what we mean by "structure": when we're judging data, what sorts of patterns are we looking for?

Often, we're looking to see, "what data is near/similar to other data?" Similar data is more likely to behave similarly, after all.

We'll come back to these ideas when we talk about clustering!

So, it might be useful to include distance between data points as a feature: how do we implement this? Well, let's do this one-by-one: we'll create a feature for the distance to a single data point p.

We start with squared distance, for smoothness reasons.

$$||p - x||^2 \tag{1}$$

This feature would *grow* as data points get further apart, though. We want to see what data is *close*: the opposite.

We could use a function like $\frac{1}{u}$. However, this would explode to infinity as distances shrink: not good.

e^{-u} is a better fit: it approaches 1 when $u = 0$, and, relatedly, it tends to drop off more smoothly and gradually than $1/u$.

Finally, we add a coefficient β to the exponent to give us more control: it will tell us how quickly our function decays with distance.

The word "decay" is used commonly for exponential decrease.

Definition 1

We define the **radial basis function**

$$f_p(x) = e^{-\beta||p-x||^2}$$

As a **feature** in the RBF feature transformation.

This transformation takes a data point p and provides a feature $f_p(x)$ that represents "**closeness**" of x to p.

Note some useful properties of this transform:

- For small distances, this feature creates a **connection** between p and our data point: representing some local "structure" of **closeness**.
- If points are far away, this effect gradually **vanishes**: points which are **far** away have very little to do with each other.
- β controls what is considered "close" and "far":
  - if β is large, points have to be very close for an effect.
  - if β is small, we have a larger "neighborhood" of points with a relevant $f_p(x)$.

Definition 2

The **radial basis functions (RBF)** transform takes each of the data points in the input, and uses it to create a set of **radial basis function** features.

Collectively, they make the **feature space**:

$$\phi(x) = \left[f_{x^{(1)}}(x), \;\; f_{x^{(2)}}(x), \;\; \cdots, \;\; f_{x^{(n)}}(x) \right]^T$$

Where:

$$f_p(x) = e^{-\beta ||p-x||^2}$$

This transform allows us to represent "closeness" within our dataset. With it, we can compare new data points to some "reference" points $x^{(i)}$.

It's often used to allow us to represent our dataset in a way that is approximate, but still useful.

This general idea is useful for problems like:

- Function approximation,
- Optimization,
- Reducing noise in signals

Reminder that "noise" just refers to anything undesired in the signal. Usually added by randomness or the environment.

This approach is not limited to the "squared distance" idea of closeness, either: if you can come up with another way to define distance, you can use the same approach.

These ways to define distance are called "distance metrics".