Explanatory Notes for 6.390

Shaunticlair Ruiz (Current TA)

Fall 2022

What λ should we choose?

There's something we ignored earlier: how do we pick the **best** value of λ? We didn't go into detail, but that value of λ will affect our algorithm's **performance**.

This λ adjusts exactly how we **learn**: how much more do we learn **specifically**, versus **generally**?

We mentioned that different λ values have different **tradeoffs**, so we need to figure out which λ value is best for our problem.

We'll need to **optimize** our λ value! Let's figure out how to go about that.

Tradeoffs: Estimation Error

High and low λ values have benefits and drawbacks. These tradeoffs can be loosely divided into **two categories**.

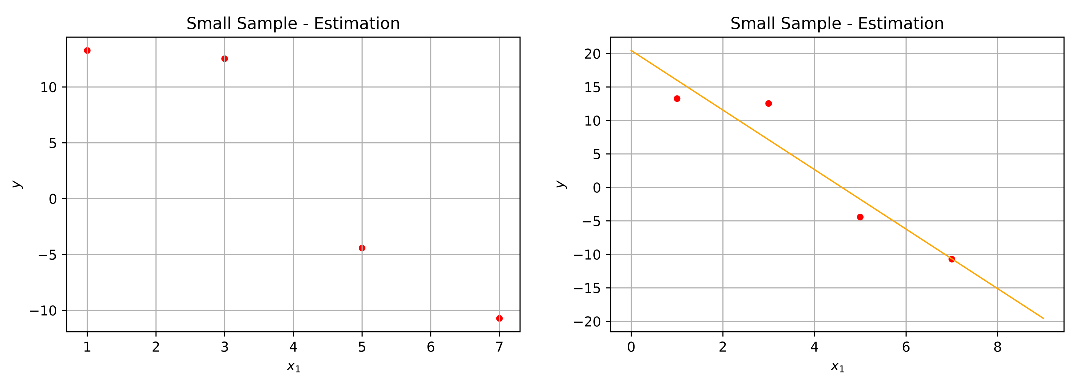
When we generalize, we're trying to avoid **estimation error**: we incorrectly guess the overall distribution we're trying to fit. We **estimate** poorly if we **generalize** poorly.

Definition 1

Estimation error is the error that results from poorly **estimating** the **solution** we're trying to find.

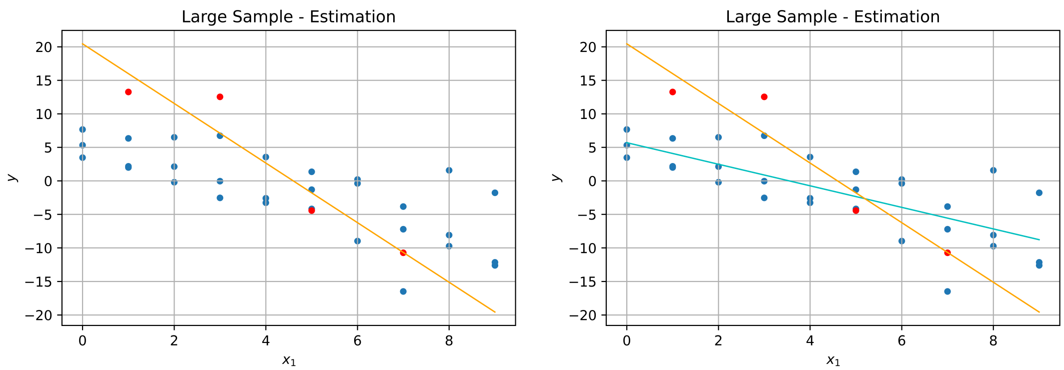
This can be caused by **overfitting**, getting a bad (**unrepresentative**) sample, or not having enough **data** to come to conclusion.

Example: Let's try a regression problem, but we'll use only 4 points to make our plot.



This is the regression solution we get based on our small dataset.

We might be suspicious. One way to reduce **estimation error** is to increase our number of data points (though this isn't always an option, or sufficient!)



Our regression from before doesn't look so good on this model... We make an updated regression, and get a more accurate result.

Clarification 2

λ doesn't lower **estimation error** in the **same way** that increasing **sample size** does, but the problem is **similar**.

Tradeoffs: Structural Error

However, not all problems are caused by estimation error: sometimes, it **isn't even possible** to get a good result - you chose the wrong **model class**.

This means the **structure** of your model is the problem, not your method of **estimation**. Thus, we call this **structural error**.

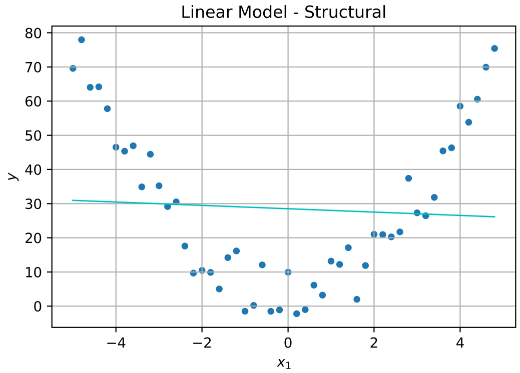
Definition 3

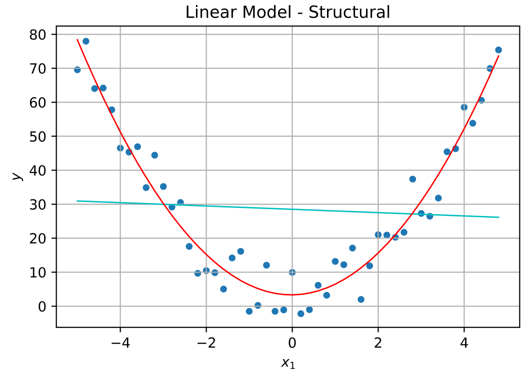
Structural error is the error that results from having the wrong **structure** for the **task** you are trying to accomplish.

This can result from the **wrong class** of model, but sometimes, your model class doesn't have the **expressiveness** it needs for a complex problem.

It can also happen if your algorithm **limits** the available models in some way, like how λ does.

Example: If the **true shape** of a distribution is a parabola x^2, there is **no** linear function $mx + b$ that can match that: this creates **structural error**.





Our **linear** model isn't able to represent a quadratic function... so, we switch to a more **expressive** model: a **quadratic** equation.

Clarification 4

Note that λ does not restrict our model class **as severely** as **switching polynomial order**, like above.

But, λ **limits** the use of larger θ, which does make it **unable** to solve some problems. So, the **structural error** problem is similar.

Remember that **expressiveness** is about how many possible models you have: if you have more models, you can solve more problems.

Tradeoffs of λ

Based on these two categories, we can discuss the tradeoffs of λ more easily.

As we mentioned, regularization **reduces** estimation error:

If we overfit to our current data, we are poorly **estimating** the distribution, because the training data may not perfectly **represent** it.

Concept 5

A **large λ** means **more regularization**: we more strongly push for a more **general** model, over a more **specific** one.

This results in...

- **Reduced** estimation error
- **Increased** structural error

However, **regularization** also **limits** the possible models we can use - those it views as less "general", it **penalizes**.

That means the scope of possible models is **smaller** - some models are no longer **acceptable**. What if the only valid solution was in that space we **restricted**? Well, then we can't **find** it.

That means there are certain **structural** limits on our model: that means that regularization

increases structural error!

Concept 6

A **small λ** means **less regularization**: we care less about a more **general** model, allowing more **specific** data to come into play.

This results in...

- **Increased** estimation error
- **Reduced** structural error