Motivation

- So far, to us, "model complexity" has essentially meant "number of features."
  The main hyperparameter we've tuned is polynomial degree. For instance, a polynomial of degree 5 has 5 features – an x, x^2, x^3, x^4, and x^5 feature.
  In the more recent example, we **manually** created several different pipelines, each of which used different combinations of features from the commute times dataset.

- Once we've created several different candidate models, we've used cross-validation to choose the one that best generalizes to unseen data.

- Another approach: **instead of manually choosing which features to include, put some constraint on the optimal parameters, $w_0^*, w_1^*, \ldots, w_d^*$.**
  This would save us time from having to think of combinations of features that might be relevant.

- Intuition: **The bigger the optimal parameters $w_0^*, w_1^*, \ldots, w_d^*$ are, the more *overfit* the model is to the training data.**
  Why?

$H(x) = 1 + 2x + 1000x^3 - 100000x^4$

$\Rightarrow$ change x a little, output changes a lot!!!!

not good!

An easier approach: `GridSearchCV`

iterates over every <u>combination</u> of hyp

- Instead of having to `for`-loop over possible hyperparameter values, we can let `sklearn` do the hard work for us, using `GridSearchCV`.

- `GridSearchCV` takes in:
    - an **un-`fit`** instance of an estimator, and
    - a **dictionary** of hyperparameter values to try,

degree → / lambda ↓	1	2	3	4	...	25
1/10	X	X	X	X	..	..
1/100	X	X	X	X	X	
1/1000		X				
⋮						

and performs k-fold cross-validation to find the **combination of hyperparameters** with the best average validation performance.

```
In [27]: from sklearn.model_selection import GridSearchCV
```

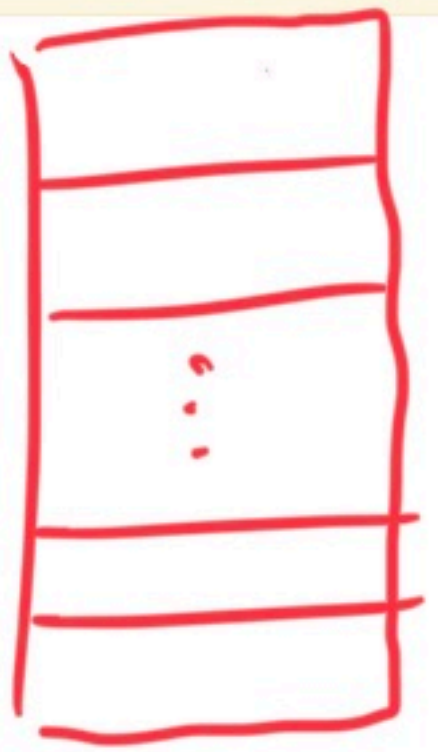
- Why do you think it's called "grid search"?

per hyperparameter: each point is used for training 9x ✓

$\Rightarrow$ 20 hyperparams, $20 \times 9 = \boxed{180}$

Question 🤔 (Answer at practicaldsc.org/q)

- Suppose you have a training dataset with 1000 rows.
- You want to decide between 20 hyperparameters for a particular model.
- To do so, you perform 10-fold cross-validation.
- **How many times is the first row in the training dataset (`X.iloc[0]`) used for training a model?**



training data

10 pieces

each piece is used for validation ONCE per hp

training 9x per hp.