OVERFITTING

AND

UNDERFITTING

One of the biggest problems

in machine learning

Remember Arthur Samuel?

"Machine learning gives computers the ability to learn without being explicitly programmed."

What does "without being explicitly programmed" mean?

Not being explicitly programmed

=

No explicit rules for how to translate the inputs into outputs

We do have a structure for how to translate inputs into outputs.

$$(w_0 * x_0) + (w_1 * x_1) = \hat{y}$$

The machine learning model

w_0, w_1 } parameters

x_0, x_1 } Inputs

$\hat{y}$ prediction

The patterns we're trying to find are complex - we use machine learning because we want to identify these implicit patterns in the training data set.

Overfitting and underfitting occur when we dissolve the structure into a set of explicit rules.

Let's look at how this happens...

Overfitting

Training data set

A_1	A_2	A_3	Target
2.5	0	1	1
4	1	0	0
10.2	0	0	1

Model

IF $A_1 = 2.5$ and $A_2 = 0$ and $A_3 = 1$
OR
$A_1 = 10.2$ and $A_2 = 0$ and $A_3 = 1$
$\Big\} \hat{y} = 1$

OR

IF $A_1 = 4$ and $A_2 = 0$ and $A_3 = 0$ $\Big\} \hat{y} = 0$

Explicit rules

Question: What is $\hat{y}$ when

$A_1 = 5.2$ and $A_2 = 1$ and $A_3 = 0$?

There's no way for this model to answer this question!

This model has just <u>memorized</u> the training data set.

Answer: The $\hat{y}$ the model returns will be unpredictable. It can be anything.

This is an extreme case of overfitting.

Another way we can think about what we want our models to do is that they be generalizeable.

Generalizeable Model = A model that predicts correctly even when it has never seen the inputs before.

After all, isn't this the point of prediction anyway?

So, if want a model that is good at prediction, you must keep from Overfitting it.

Overfitting occurs when:

- you have too little training data (# rows or # attributes)

AND

- your model is too complex

OR

- You let your model spend too much time with the training data.

Underfitting

The same training data set we had before

A_1	A_2	A_3	Target
2.5	0	1	1
4	1	0	0
10.2	0	0	1

Model

IF $A_1 = 2.5$ OR $A_1 = 4$ or $A_1 = 10.2$

OR

$A_2 = 0$ OR $A_2 = 1$ or $A_2 = 0$

OR

$A_3 = 1$ OR $A_3 = 0$ OR $A_3 = 0$

} Predict $\hat{y} = 0$

OR

IF A_1, A_2, A_3 are any other values, predict $\hat{y} = 0$

Question: What is $\hat{y}$ when

$A_1 = 5.2$ and $A_2 = 1$ and $A_3 = 0$?

Question: What is $\hat{y}$ when

$A_1 = 6.2$ and $A_2 = 1$ and $A_3 = 1$?

Question: What is $\hat{y}$ when

$A_1 = 2.5$ and $A_2 = 0$ and $A_3 = 1$?

...

Starting to see a pattern here?

This model just predicts the same thing no matter what inputs you provide it.

The model hasn't memorized your data set - rather it's ignored it.

This is an extreme case of of a model that is underfit.

Underfitting occurs when:

- you have a lot of training data
  (# rows or # attributes)

  AND

- your model is too simple

Summary

- We use machine learning models to predict things.
- To predict well, a model needs to be generalizable. It needs to go beyond inputs that it has seen to inputs that it has never seen. Models should not memorize inputs.
- To predict well, a model must be able to make use of the training data inputs. Models should not ignore inputs.

(Contd.)

■ The 3 factors that play a role in overfitting and underfitting.

1) Amount of training data

2) Complexity of the model

3) Amount of time the model spends with the training data

We have some control over all 3 factors.