Bias in a machine learning model is not the same thing as bias in an estimator.

From CS229:

The learning algorithm has two sources of information:

* The model’s set of assumptions, which is also known as inductive bias. This is the information that comes from you, the data scientist. The stronger the assumptions, the higher the inductive bias. If the model’s inductive bias is wrong – that is, you’ve made wrong assumptions about the data – then the model will underfit the data. (Inductive bias is a completely different concept from bias in an estimator.)
* The training data

Points that are roughly linear (with noise). An example of inductive bias is picking a linear model. Less inductive bias: pick a high-order polynomial that will fit the points exactly.

# Inductive bias

## Inductive reasoning example

In Switzerland, you come across a spotted cow with a cowbell. You may assume that all spotted cows in Switzerland have a cowbell. This is an example of inductive reasoning: it starts with an observation (a cow with spots and a cowbell) and leads to a possible generalization hypothesis (all cows with spots have a cowbell).

It’s possible to draw/induce other hypotheses based on the same observation.

* There are cows in Switzerland
* All cows have cowbells
* There are only cows in Switzerland
* Etc.

This is an important property of inductive reasoning: based on an observation, you can induce any number of hypotheses, and some of them can be false.

How do you choose a single hypothesis? One option is to choose the simplest hypothesis: “There are cows in Switzerland.” This approach is called Occam’s razor: the simplest consistent hypothesis is the best one.

Side note: Inductive reasoning provides conclusions that are at best probable, given the evidence provided. Deductive reasoning generates a conclusion that is certain, given the premises are correct.

## Machine learning

In most machine learning tasks, we have a set of observations (training data), and we want to create a generalization (hypothesis/model) based on them. We want this generalization to be valid for new, unseen samples.

In other words, based on a subset of samples, we want to induce a general rule that applies to the whole population.

Like in the cow example, we can have many different hypotheses (models). Each model can fit the training data but provides significantly different results on unseen data.

The inductive bias (also known as learning bias) of a learning algorithm is the set of assumptions that the learner uses to predict outputs of given inputs that it has not encountered. Inductive bias is anything which makes the algorithm learn one pattern instead of another pattern (e.g. step-functions in decision trees instead of continuous functions in linear regression models).

Making weaker assumptions leads to a more robust model, but for small datasets, making stronger (and correct) assumptions will lead to better results. Larger datasets naturally give more information, so you can get away with making weaker assumptions. This is a general principle/tradeoff of machine learning.

Algorithm has two sources of information: the model assumptions (information that you give), and the data. When you have less data, your skill making the model assumptions is much more important.

Quick aside: a polynomial of order is uniquely defined by points on that polynomial, e.g. 2 points define a line, 3 points define a quadratic, and so on. For , the points cannot be collinear (on the same line).

This polynomial has unknowns, so you need points to specify a system of linearly independent equations to solve for the coefficients.

A white board with writing on it

AI-generated content may be incorrect.

High bias (underfitting): Does not capture the trend of the training data. The learning algorithm has very strong preconceptions (inductive bias) about the input-output relationship of the data that aren’t true. Reduce bias to allow the model to learn from the data (remember from the GDA discussion that the model has two sources of information – the assumptions and the data).

High variance: Captures the training data trend too closely; captures the noise. A different set of training examples will cause the model predictions to vary wildly. Increase bias to prevent the model from adhering too closely to the data.

Choosing the model and its hyperparameters often benefits from domain knowledge.

# References

* <https://medium.com/data-science/the-inductive-bias-of-ml-models-and-why-you-should-care-about-it-979fe02a1a56>
* CS229 notes
* <https://en.wikipedia.org/wiki/Inductive_bias>
* <https://en.wikipedia.org/wiki/Inductive_reasoning>