

Making Real AI - Series

3. 100% Training Accuracy without Overfitting

(Shaka) Shih-Chia Chen

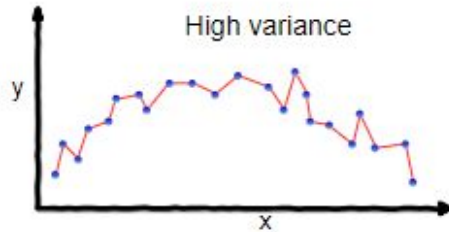
Founder/CEO
www.libgirl.com



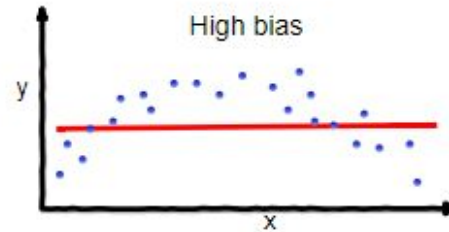
Background: Underfitting VS. Overfitting

Overfitting: The production of an analysis which corresponds too closely.....to a particular set of data, and may therefore fail to predict future observations reliably. (*Overfitting | Meaning of Overfitting by Lexico, n.d.*)

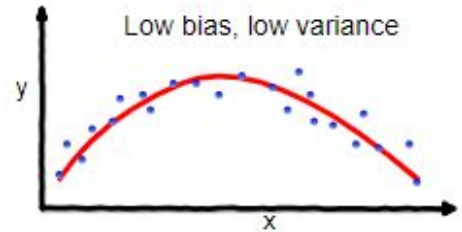
Underfitting: Underfitting occurs when a statistical model cannot adequately capture the underlying structure of the data. (Wikipedia Contributors, 2019)



overfitting



underfitting



Good balance

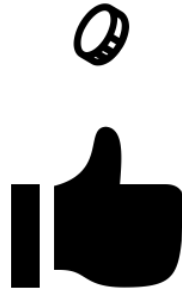
Thesis:
**No overfitting problem in context sensitive
machine learning system.**

Overfitting Happens If

The essence of overfitting is to have unknowingly extracted some of the **residual variation (i.e. the noise)** as if that variation represented underlying model structure.

(Burnham & David Raymond Anderson, 2002/2010)

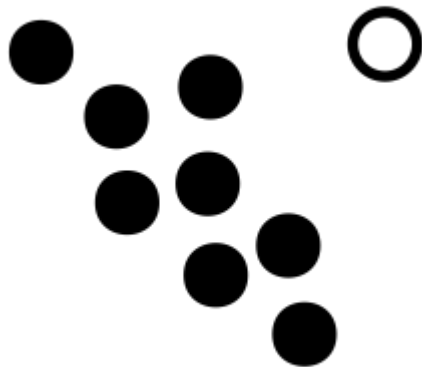
But is there any real noise?



Flipping coin is random
only if we have no clues about it

Data Has Its Context

The seemingly 'noise' has its context in which it is generated.



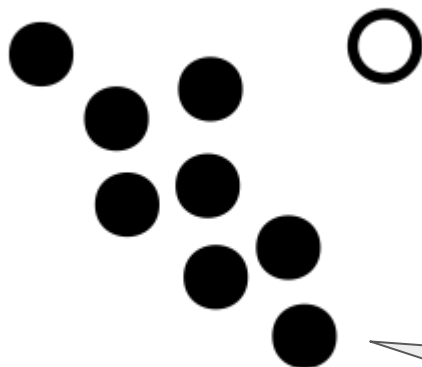
Maybe a person mis-presses the sensor.

Maybe an earthquake happens.

Maybe a lightning hits the laboratory.

Sensitive to the Right Context

As long as the model can do inference based on the right context, fitting all the training data won't tradeoff your future accuracy.



Use this outlier information if only if the same context re-occurs !

Other contexts, use regularity learnt here.

Conclusion

Enough data with enough contextual information

+

Very large model to learn all the regularities according to their contexts

=

100% Training Accuracy without Overfitting.

Related Works

- Reconciling modern machine learning practice and the bias-variance trade-off (Belkin et al., 2019)
- Deep double descent: where bigger models and more data hurt (Nakkiran et al., 2019)

Next Slides

It's impossible to completely collect contextual information at the very beginning, but we can find more later on.

Thus we need,

4. An example way to append new contextual information into old pieces of data.

Reference

- Singh, S. (2018, May 21). *Understanding the Bias-Variance Tradeoff*. Medium; Towards Data Science. <https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229>
- Wikipedia Contributors. (2019, February 23). *Overfitting*. Wikipedia; Wikimedia Foundation. <https://en.wikipedia.org/wiki/Overfitting>
- Burnham, K. P., & David Raymond Anderson. (2010). *Model selection and multimodel inference : a practical information-theoretic approach* (2nd ed.). Springer. (Original work published 2002)
- Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32), 15849–15854. <https://doi.org/10.1073/pnas.1903070116>
- Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Openai, I. (2019). *DEEP DOUBLE DESCENT: WHERE BIGGER MODELS AND MORE DATA HURT*. <https://arxiv.org/pdf/1912.02292.pdf>