

Double descent

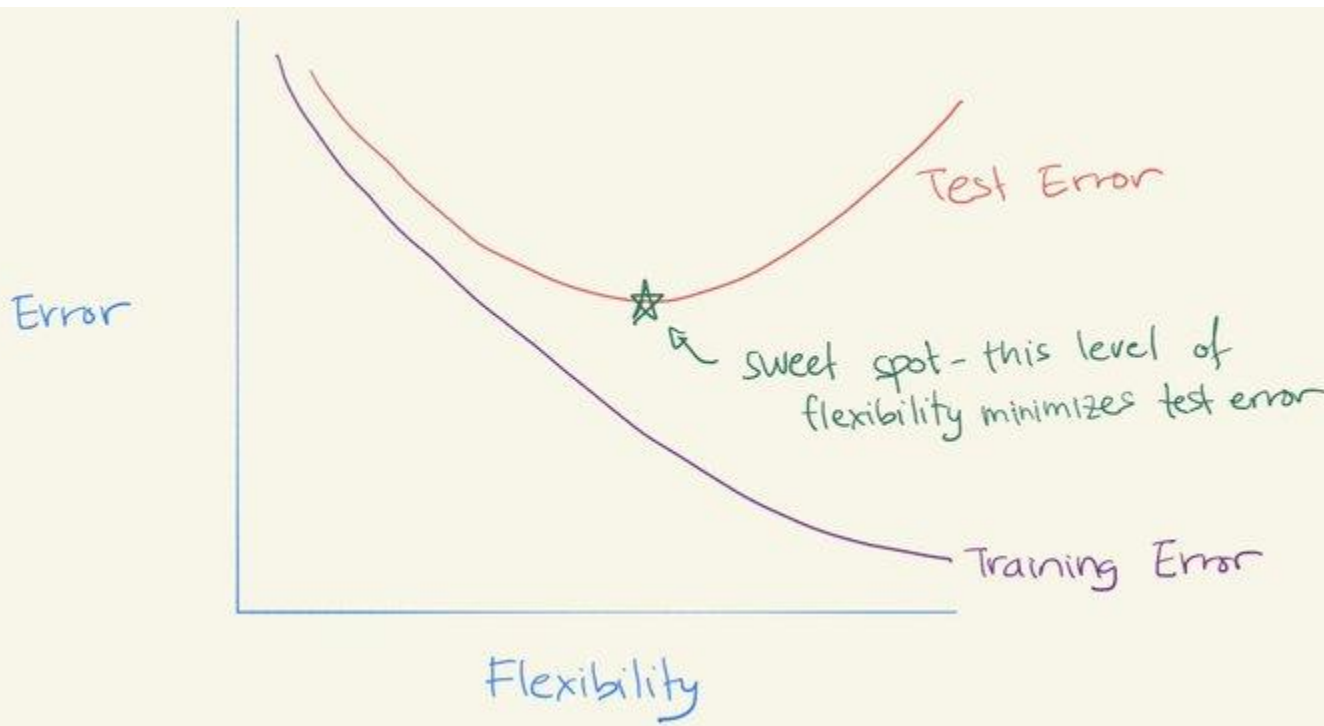
Is the bias-variance trade-off still valid in the era of deep learning?

Filippo Biscarini
Senior Scientist
CNR, Milan (Italy)

Nelson Nazzicari
Research fellow
CREA, Lodi (Italy)



Bias-variance trade-off

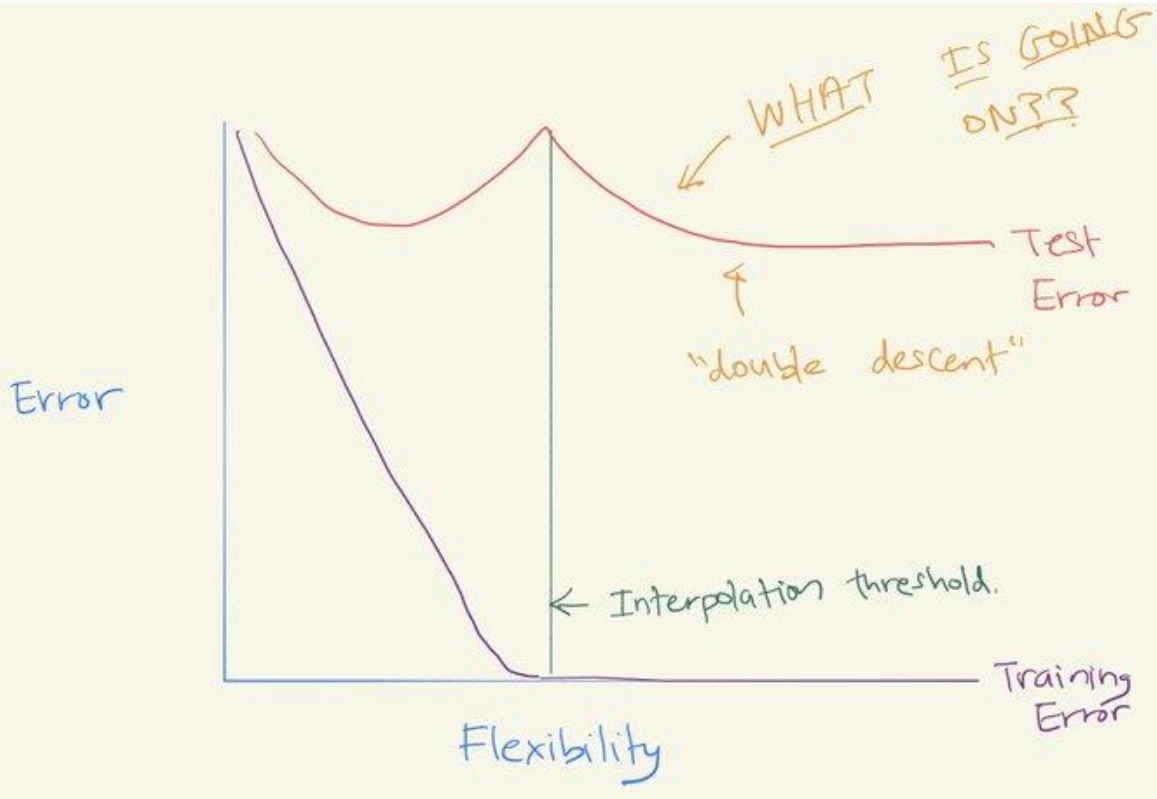


- $\text{Err} = \text{var} + \text{bias}^2 + \text{irred_err}$
- fine-tuning of hyperparameters is essentially trading around variance and bias
- the aim is to get the lowest possible prediction error (in test data)
- flexibility is the same as model complexity

From Daniela Witten



Double descent



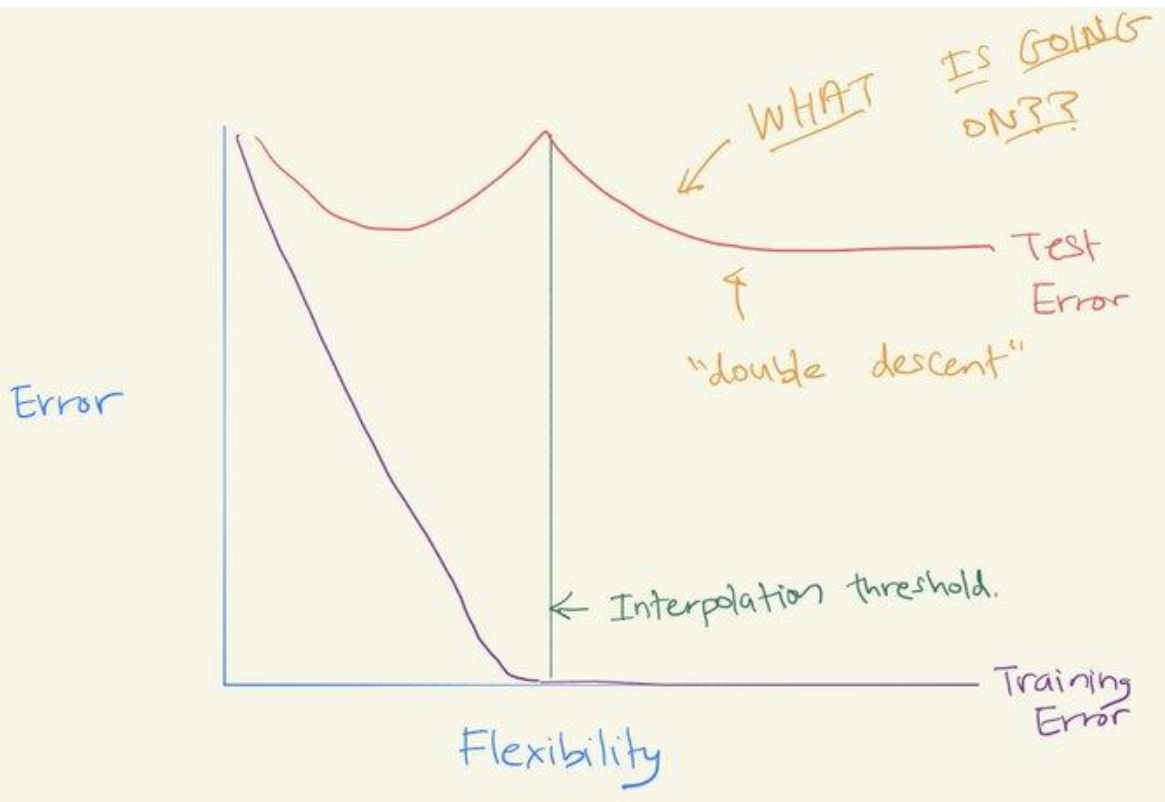
double descent

deep learning → increasingly complex models that give a second descent of the test error

remember: DL models can have millions of parameters



Double descent



double descent

deep learning → increasingly complex models that give a second descent of the test error

remember: DL models can have millions of parameters

“scientific” questions

is the bias-variance trade-off not true?

is deep learning ‘magic’?



A simplified insight

polynomial regression

$$y = \beta_0 + \beta_{i1}X_i + \beta_{i2}x_i^2 + \beta_{i3}x_i^3 + \dots + \beta_{id}x_i^d + \epsilon$$

- many parameters “p”
- highly non-linear functions

} similar to deep learning



A simplified insight

polynomial regression

$$y = \beta_0 + \beta_{i1}X_i + \beta_{i2}x_i^2 + \beta_{i3}x_i^3 + \dots + \beta_{id}x_i^d + \epsilon$$

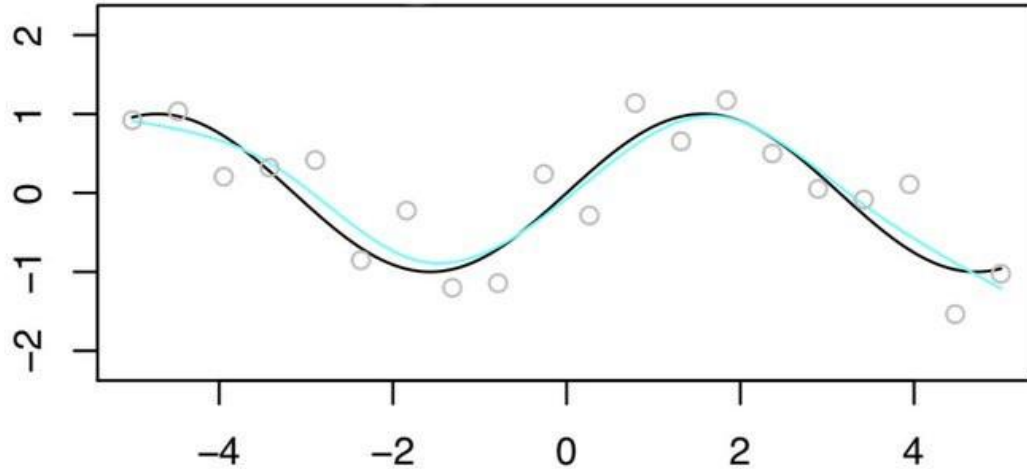
- $n = 20$ (sample size)
- increasing n. of parameters p
- solve by least square



A simplified insight

polynomial regression: $n = 20$, $p = 6$

6P



good fit! (black line: true function; blue line: fitted function)

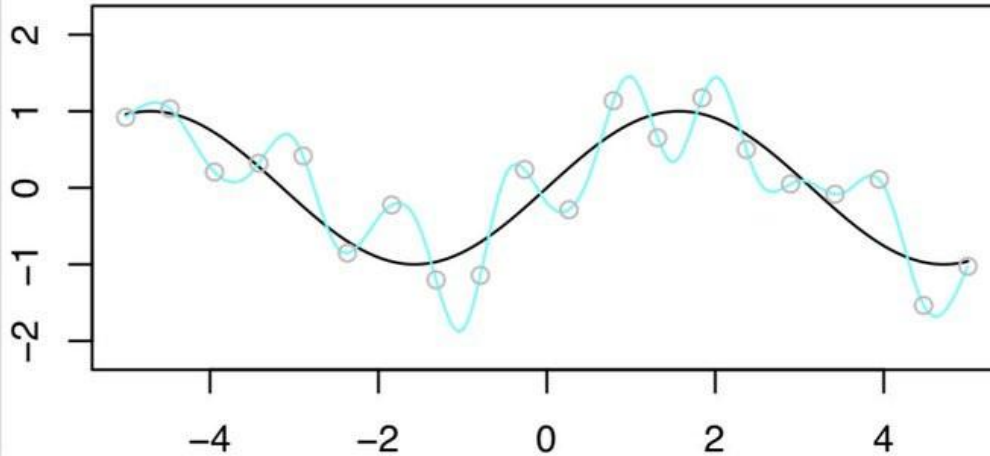
From Daniela Witten



A simplified insight

polynomial regression: $n = 20$, $p = 20$

20 P



- only one least-square fit
- **zero training error**
- no generalization → **very high test error**
- complex model, with high variance and low bias

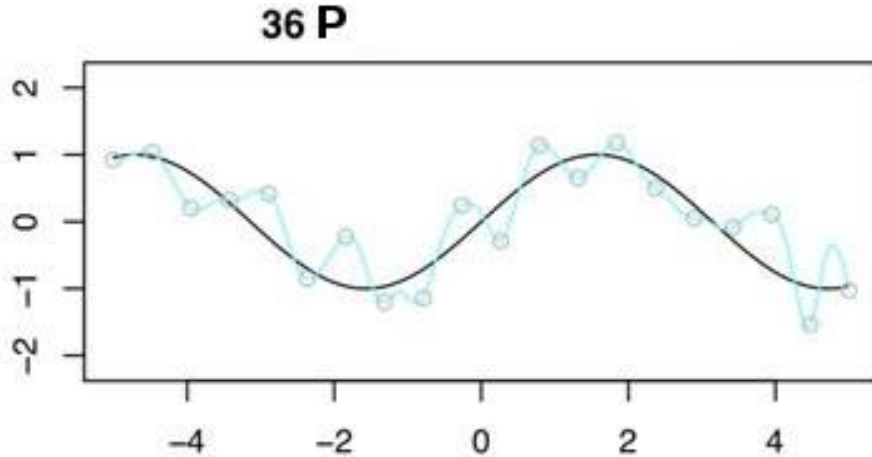
bias-variance trade-off

From Daniela Witten



A simplified insight

polynomial regression: $n = 20$, $p = 36 \rightarrow p > n$



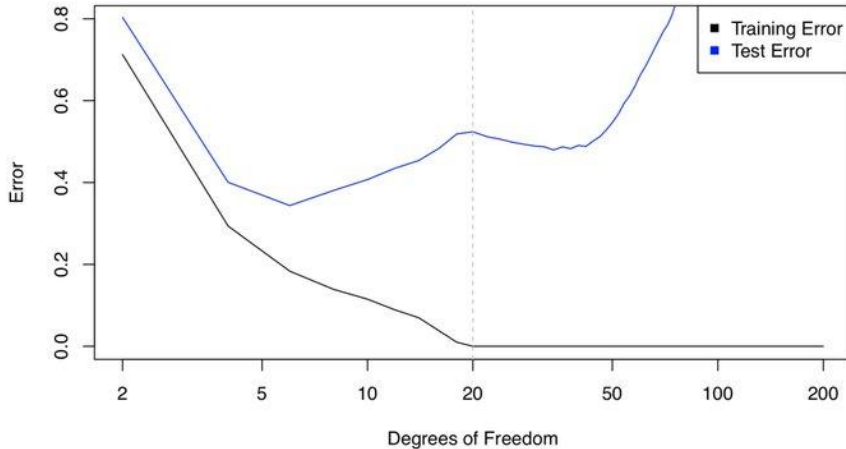
looks bad, but not worse than
model with $p = 20$

From Daniela Witten



A simplified insight

polynomial regression: $n = 20$, $p = 36 \rightarrow p > n$



- increasing p , beyond n
- double descent: test error decreases briefly when $p > n$ (then goes up again)

is the bias-variance trade-off not true?

From Daniela Witten



To recap

- **double descent is a real thing** that happens
- it is not magic!
- the **bias-variance trade-off still holds** and helps explain double descent

All this came from a discussion led by **Daniela Witten**:

<https://www.biostat.washington.edu/people/daniela-witten>

Check her material for more details and a less simplistic explanation

