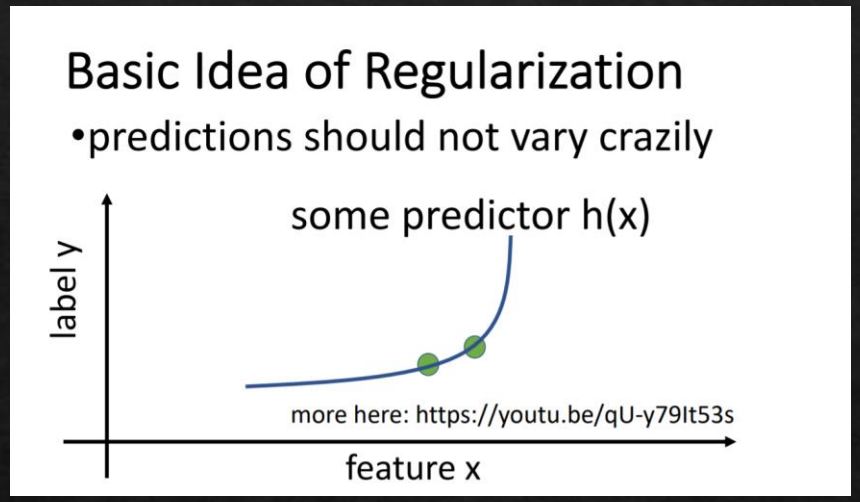
Regularization parameter C

Yongbing Anita Tao-Topinoja

What regularization parameter C does in scikit-learn LogisticRegression?

- What regularization does?
  - “Smooth model selection”, (model = Some Hypothesis Space)

- What does regularization parameter “C” represent in scikit-learn LogisticRegression?
  - “C” represents the inverse of regularization strength, which must always be a positive float.

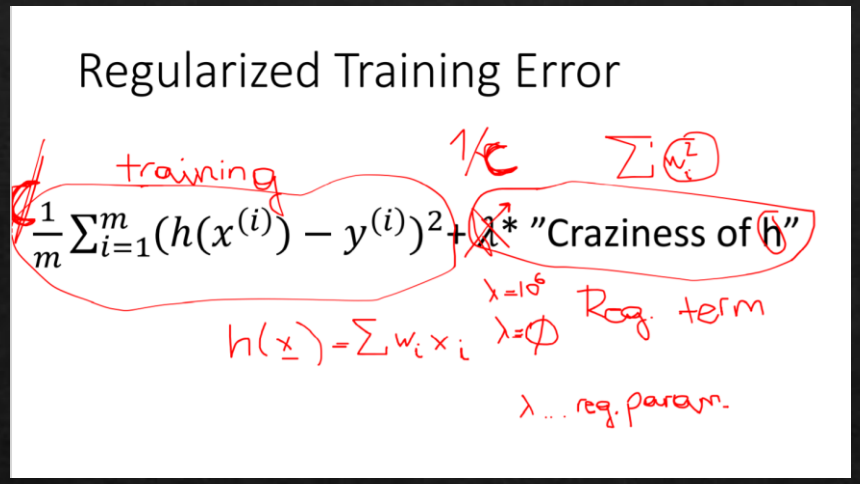


What regularization parameter C does in scikit-learn LogisticRegression?

- To control the regularization strength, e.g. to control predictor h(x) not changing sharply.
- We could think that regularization term is as adding variance to our model.

Source: CS-EJ3211 Machine Learning with Python, Aalto University

How to choose appropriate C?



- In scikit-learn LogisticRegression,
  - the default of C is 1.0,
  - smaller values of C specify stronger regularization.
- If our model suffers from high variance,

  e.g. for overfitting, smaller "C" is needed.
- On the other hand, too much variance will result in underfitting. In that case, "C" should not be too small.

Source: CS-EJ3211 Machine Learning with Python, Aalto University