

# Regularization parameter $C$

Yongbing Anita Tao-Topinoja

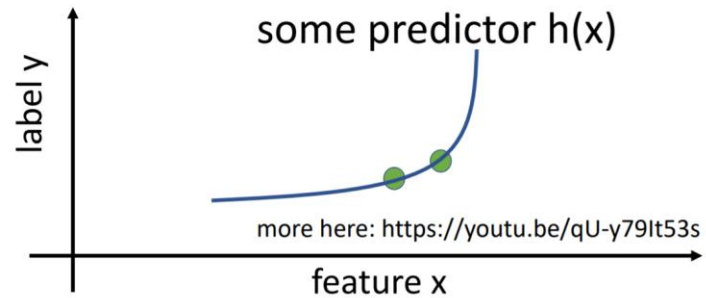
# What regularization parameter C does in scikit-learn LogisticRegression?

- ◇ 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?

## Basic Idea of Regularization

- predictions should not vary crazily



- ◇ 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?

## Regularized Training Error

$$\frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{i=1}^n w_i^2$$

Handwritten notes in red:

- $\frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2$  is labeled "training".
- $\lambda$  is labeled "Craziness of h".
- $\lambda = 10^6$  and  $\lambda = 0$  are written near the regularization term.
- $\lambda \dots \text{reg. param.}$  is written below the regularization term.
- $h(x) = \sum w_i x_i$  is written below the first term.

- ◇ 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.