

# Content

## 0. Introduction

## 1. Regression

1.1 Multivariate Linear Regression (curve fitting)

1.2 Regularization (Lagrange multiplier)

1.3 Logistic Regression (Fermi-Dirac distribution)

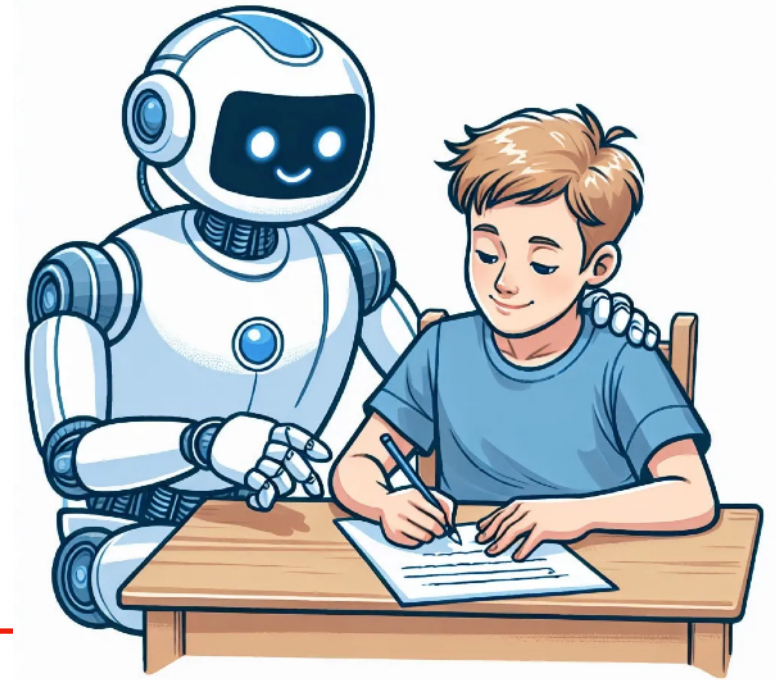
1.4 Support Vector Machine (high-school geometry)

## 2. Dimensionality Reduction/feature extraction

2.1 Principal Component Analysis (order parameters)

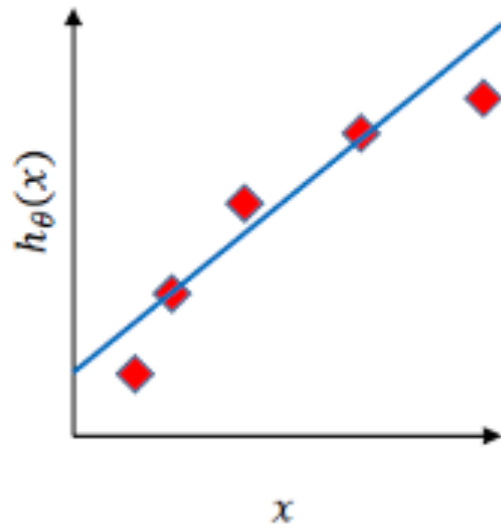
2.2 Recommender Systems

2.3 Clustering (phase transition)



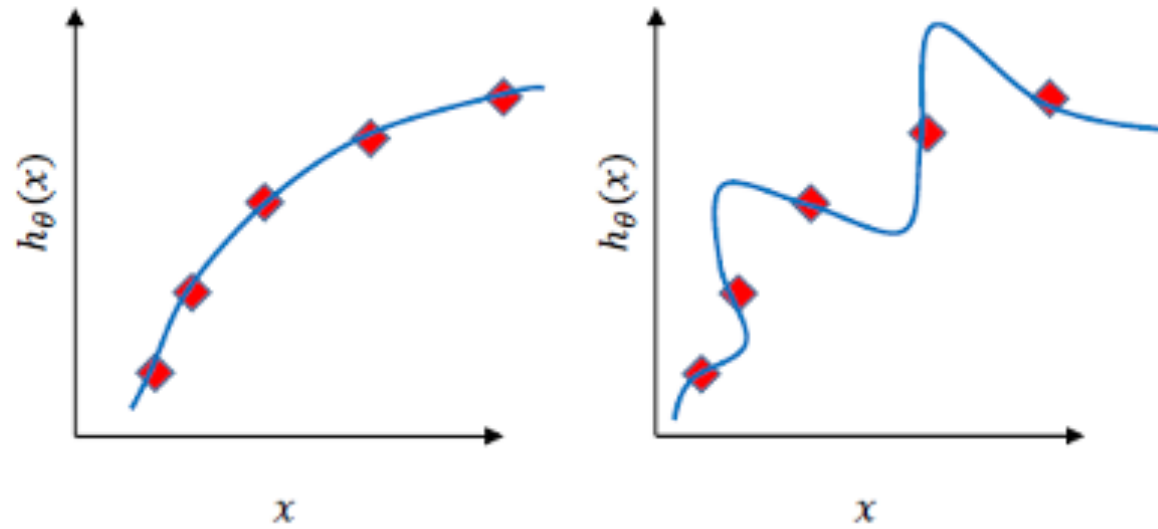
# The quality of fitting

Underfitting (high bias)



High bias

Overfitting (high variance)

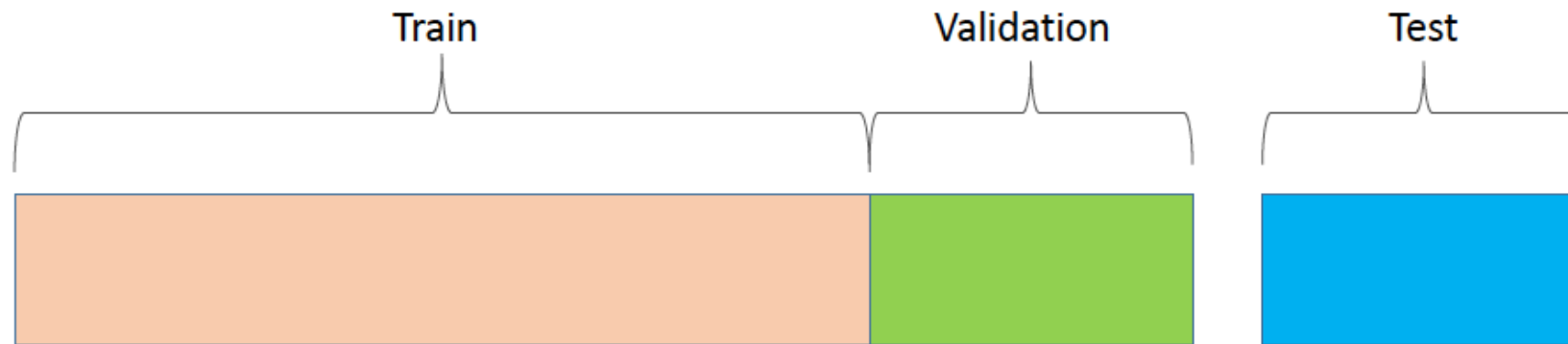


High variance

Polynomial regression

## Forecastability

$$h_{\theta}^{[d]}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_d x^d = \theta^T x$$

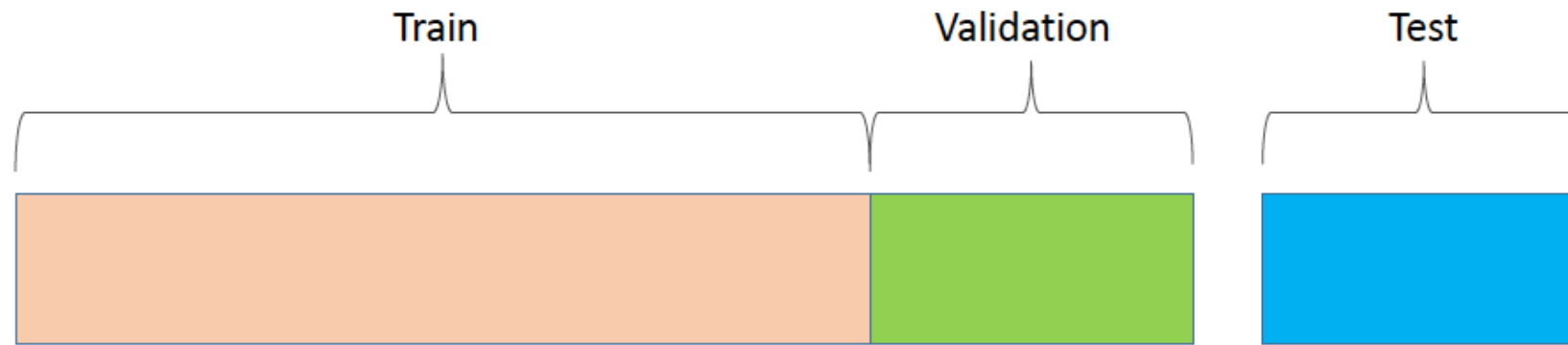


$$J_{\text{train}}(\theta) = \frac{1}{2M_{\text{train}}} \sum_{i_{\text{train}}=1}^{M_{\text{train}}} (y^{(i_{\text{train}})} - h_{\theta}(x))^2$$

$$J_{\text{cv}}(\theta) = \frac{1}{2M_{\text{cv}}} \sum_{i_{\text{cv}}=1}^{M_{\text{cv}}} (y^{(i_{\text{cv}})} - h_{\theta}(x))^2$$

$$J_{\text{test}}(\theta) = \frac{1}{2M_{\text{test}}} \sum_{i_{\text{test}}=1}^{M_{\text{test}}} (y^{(i_{\text{test}})} - h_{\theta}(x))^2$$

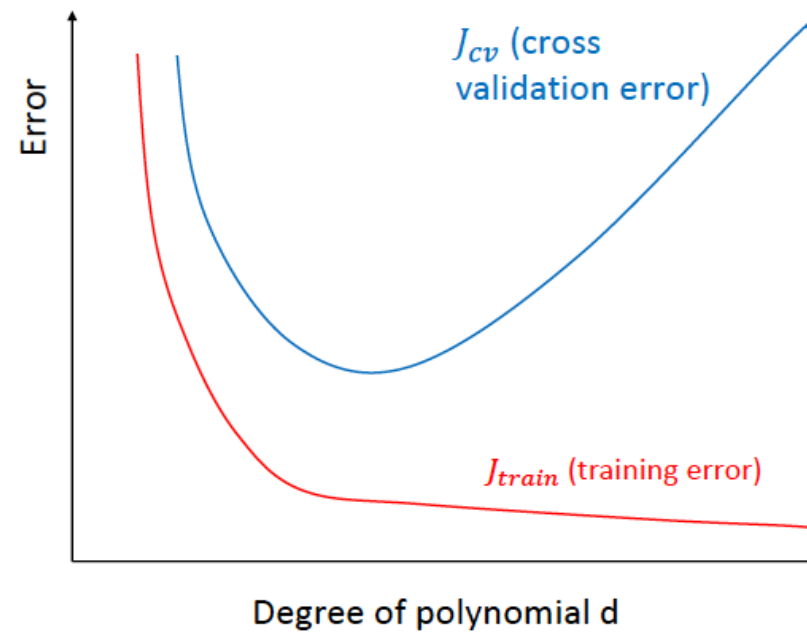
# Learning Curves and Regularization



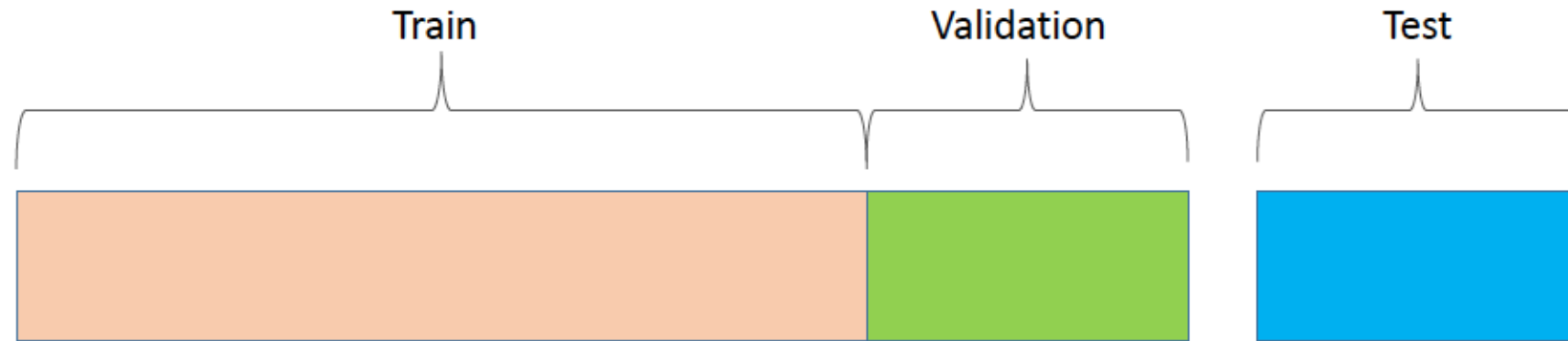
**Forecastability**

$$h_{\theta}^{[d]}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_d x^d = \theta^T x \quad \text{Polynomial regression}$$

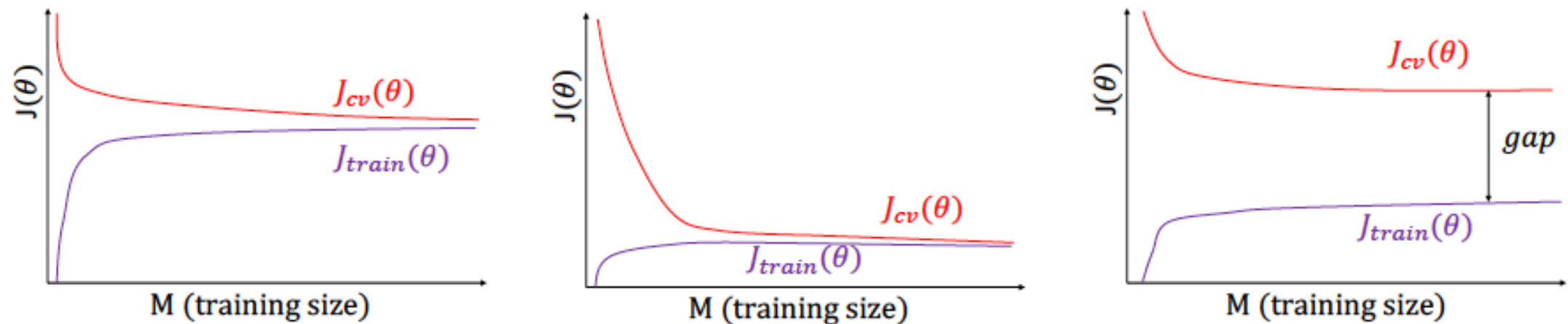
**Model Selection**



# Learning Curves and Regularization



## Learning curves



## Regularization

$$J(\theta) = \frac{1}{2M} \left[ \sum_{i=1}^M (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^N \theta_j^2 \right] \quad \lambda \geq 0$$