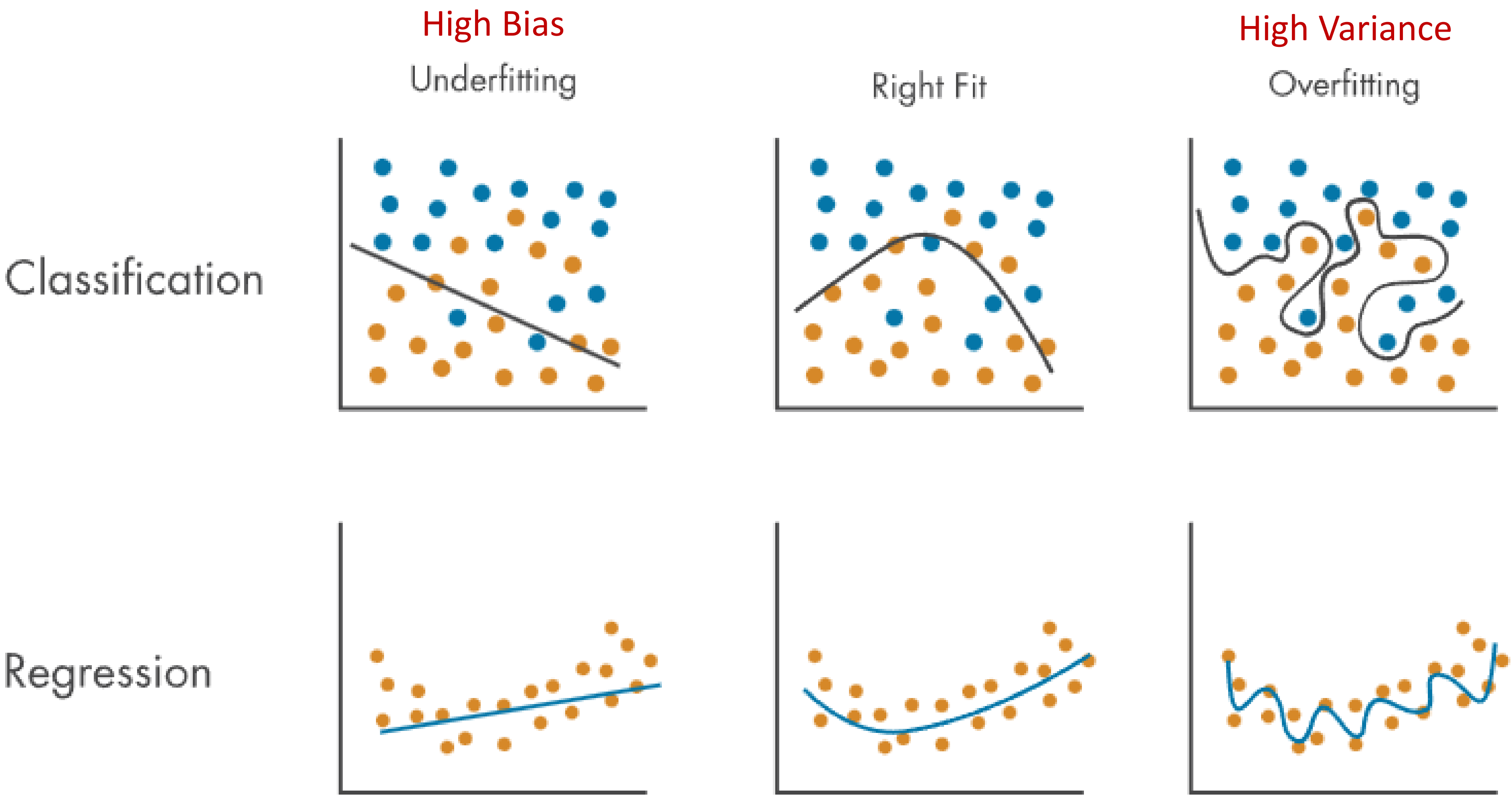
ELTE | FACULTY OF INFORMATICS

REGULARIZATION

Machine Learning Course
Balázs Nagy, PhD

ELTE | IK
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

Fit problems



Fit problems

- **Underfit:**
  - Model is too simple, need more features
- **Right fit:**
  - Nothing to do, model is good
- **Overfitting**:
  - If we have too many features, the learned hypothesis may fit the training set very well, but fail to generalize to new examples (predict on new examples)
  - Model is too complex, has too many features

Fit problems

- **Underfit:**
  - Model is too simple, need more features
- **Right fit:**
  - Nothing to do, model is good
- **Overfitting**:
  - If we have too many features, the learned hypothesis may fit the training set very well, but fail to generalize to new examples (predict on new examples)
  - Model is too complex, has too many features

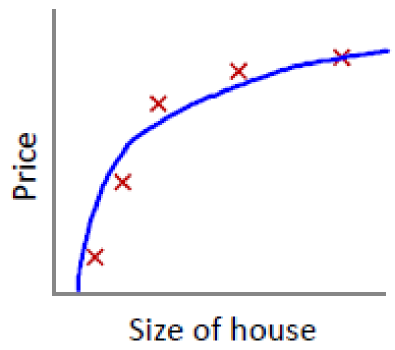
How to prevent overfit?

Prevent overfitting

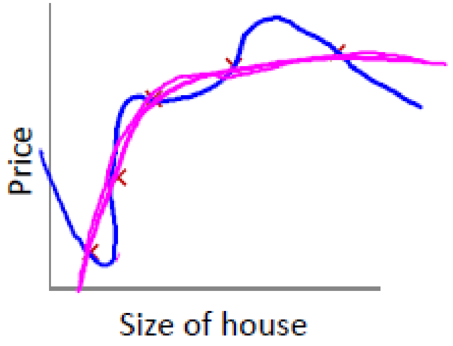
- Reduce number of features
  - Manually select which features to keep
  - Model selection algorithm
- Regularization
  - Keep all the features, but reduce magnitude / values of parameters Θ
  - Works well when we have a lot of features, each of which contributes a bit to predicting y

Regularization – Linear Regression

- Suppose we penalize high rank element and make Θ_3, Θ_4 really small



$\theta_0 + \theta_1 x + \theta_2 x^2$



$\theta_0 + \theta_1 x + \theta_2 x^2 + \cancel{\theta_3 x^3} + \cancel{\theta_4 x^4}$

$$min_\theta \ \frac{1}{2m} \ \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2$$

NOTE: Θ_0 is **not** penalized

- The λ is the regularization parameter. It determines how much the costs of our theta parameters are inflated.

Regularized Gradient descent

Repeat {

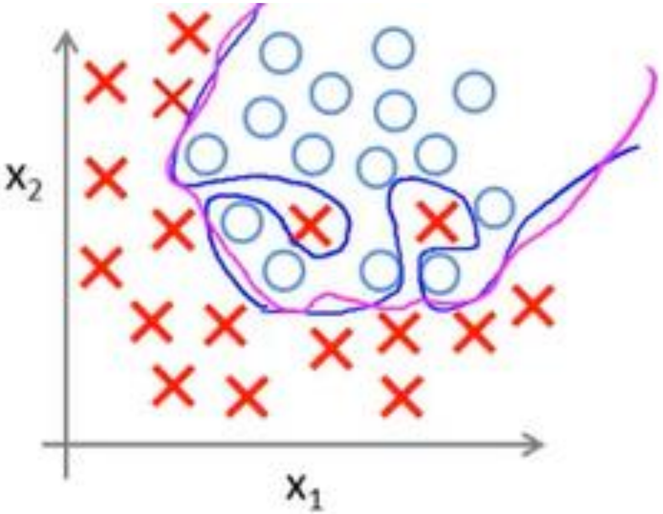
$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$\theta_j := \theta_j - \alpha \left[\left(\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \right) + \frac{\lambda}{m} \theta_j \right] \qquad j \in \{1, 2 \ldots n\}$$

}

$$\theta_j := \theta_j (1 - \alpha \frac{\lambda}{m}) - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Regularization – Logistic Regression



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \dots)$$

Cost function:

$$J(\theta) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log\left(1 - h_\theta(x^{(i)})\right)\right]$$

$$+ \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2 \qquad \boxed{\theta_1, \theta_2, \dots, \theta_n}$$

ELTE | FACULTY OF INFORMATICS

Thank you for your attention!