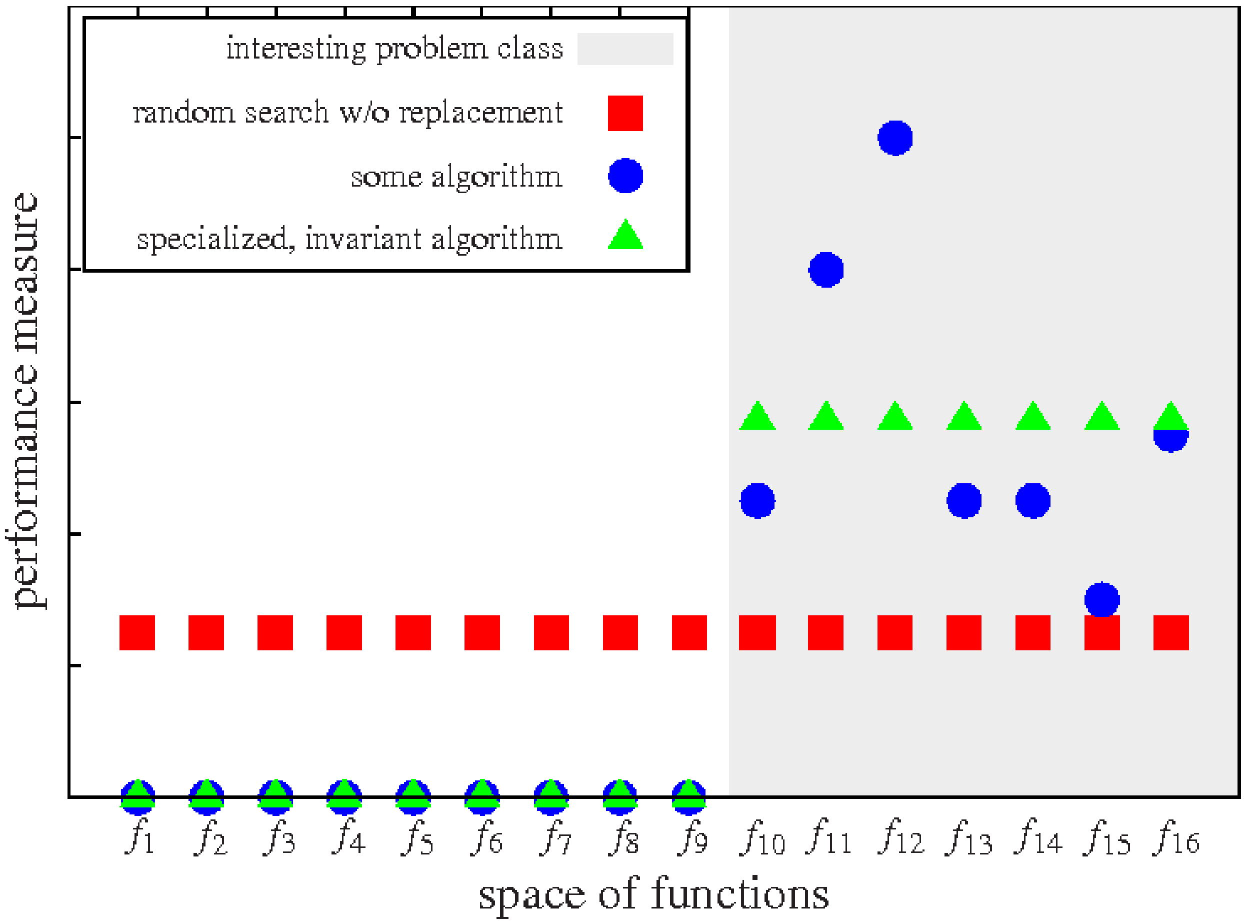
Chapter 3.1 ~ 3.3

한상유

How to choice classifier model

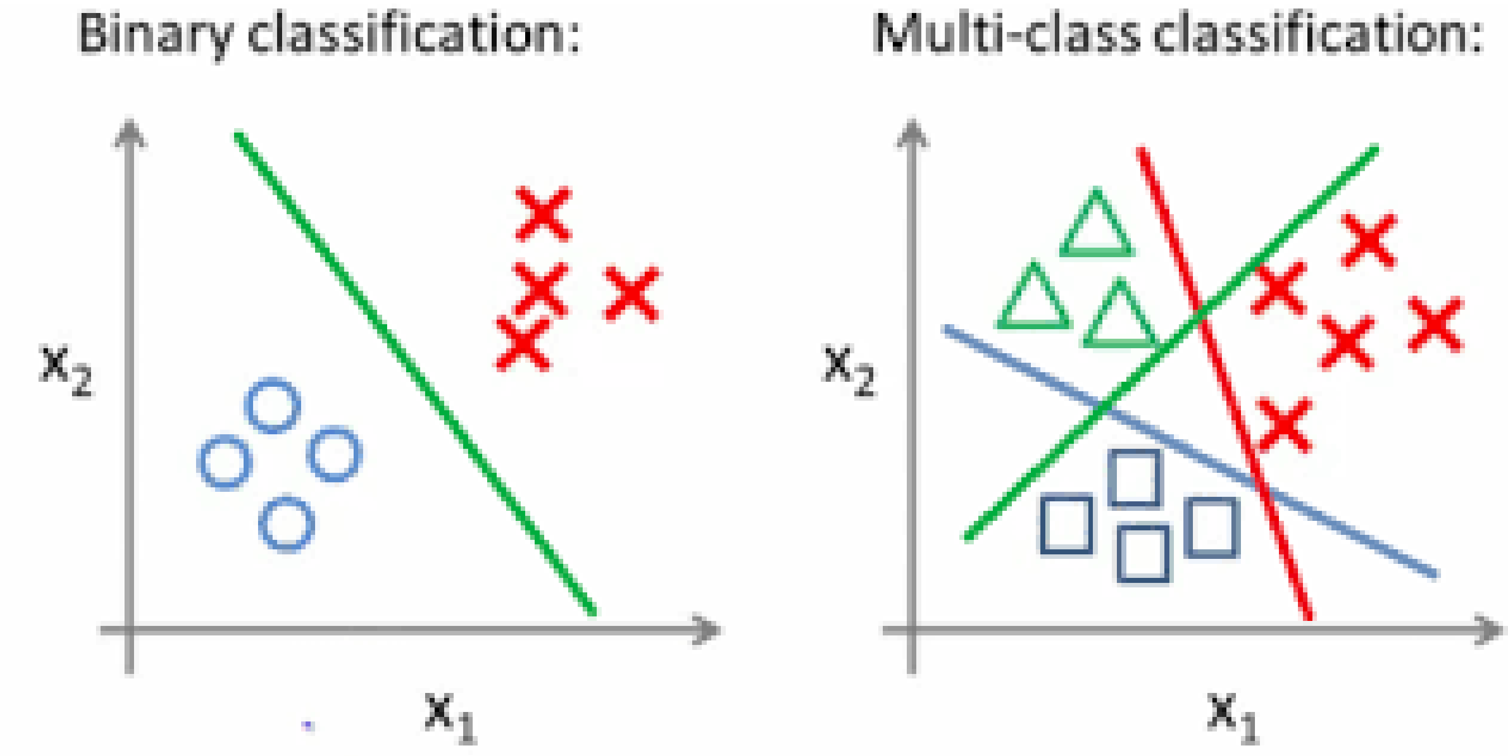


The no free lunch theorem

Machine learning algorithm training step

1. Dataset building and feature selection
2. Performance measure definition
3. Classifier and optimization algorithm selection
4. Evaluation model performance
5. Model tuning

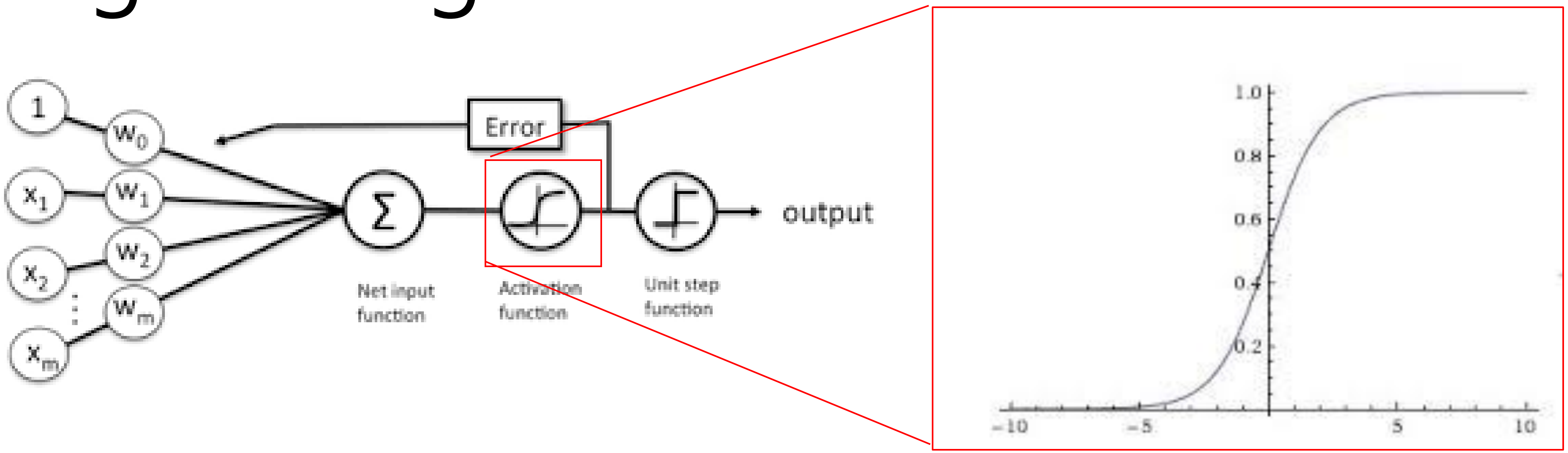
How to make multiple-class classifier



Use many classifier
(One-versus-rest)

Logistic regression

Logistic regression



- Activation function output: probability (0 ~ 1) --> Regression
- Unit step function output: class (-1, 1) --> Logistic

Logistic regression optimization

- Maximum likelyhood -> argmax(y|z)

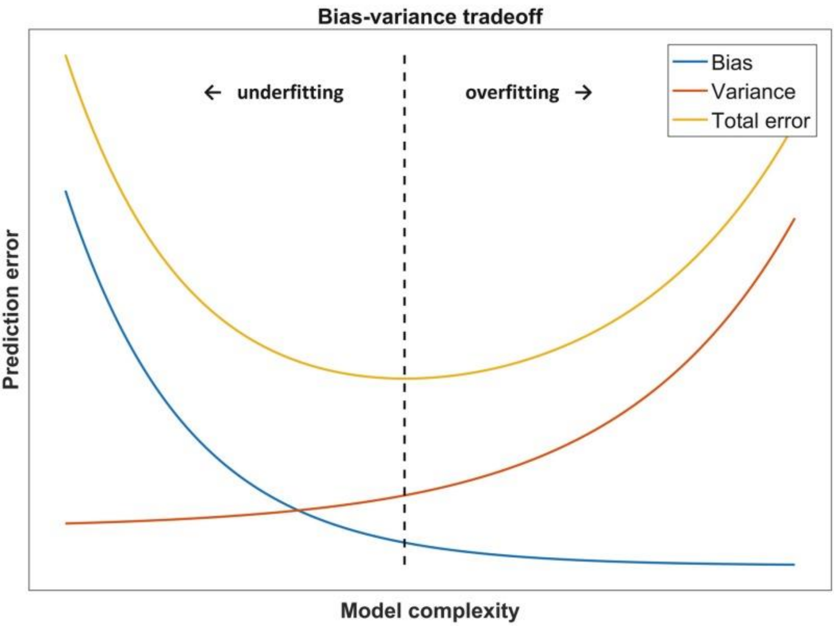
$$l(z) = -\log\Big(\prod_{i}^{m} \mathbb{P}(y_i|z_i)\Big)$$

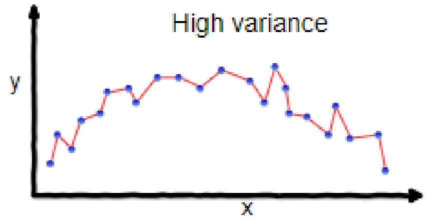
- **Log cross-entropy form**

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

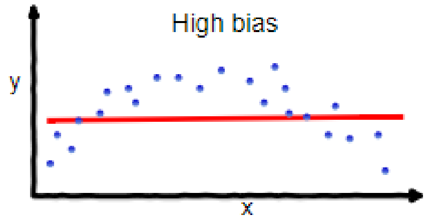
Regularization

Bias-variance trade-off

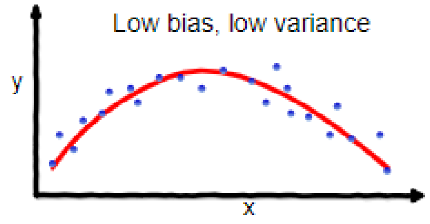




overfitting

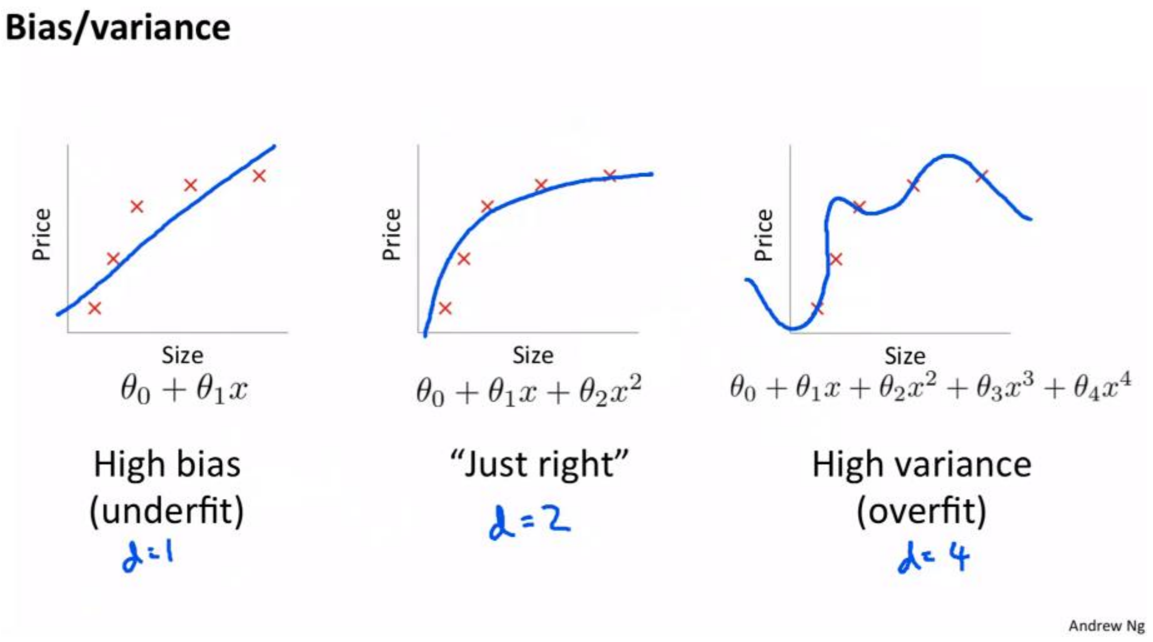


underfitting



Good balance

How to prevent overfitting



- Use a regularization (weight decay)

Regularization formula

L2: $\frac{\lambda}{2}\|\mathbf{w}\|^2 = \frac{\lambda}{2}\sum_{j=1}^{m} w_j^2$

$$J(\mathbf{w}) = \sum_{i=1}^{n}\left[-y^{(i)}\log\left(\phi(z^{(i)})\right) - \left(1-y^{(i)}\right)\log\left(1-\phi(z^{(i)})\right)\right] + \frac{\lambda}{2}\|\mathbf{w}\|^2$$