Statistical Machine Learning

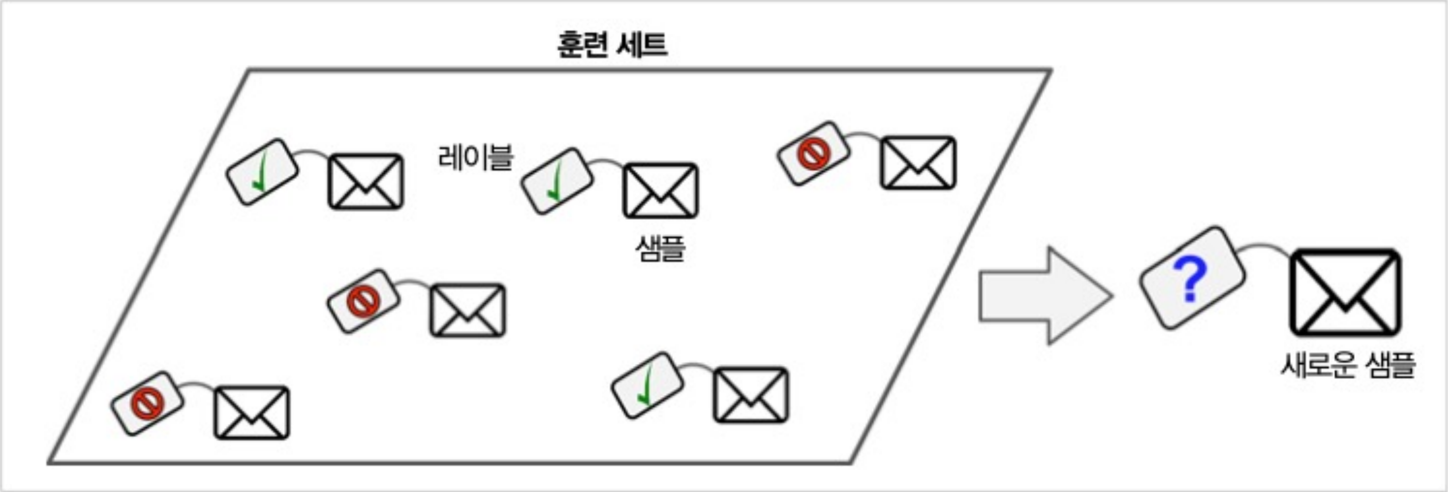
2주차

담당: 15기 김지호

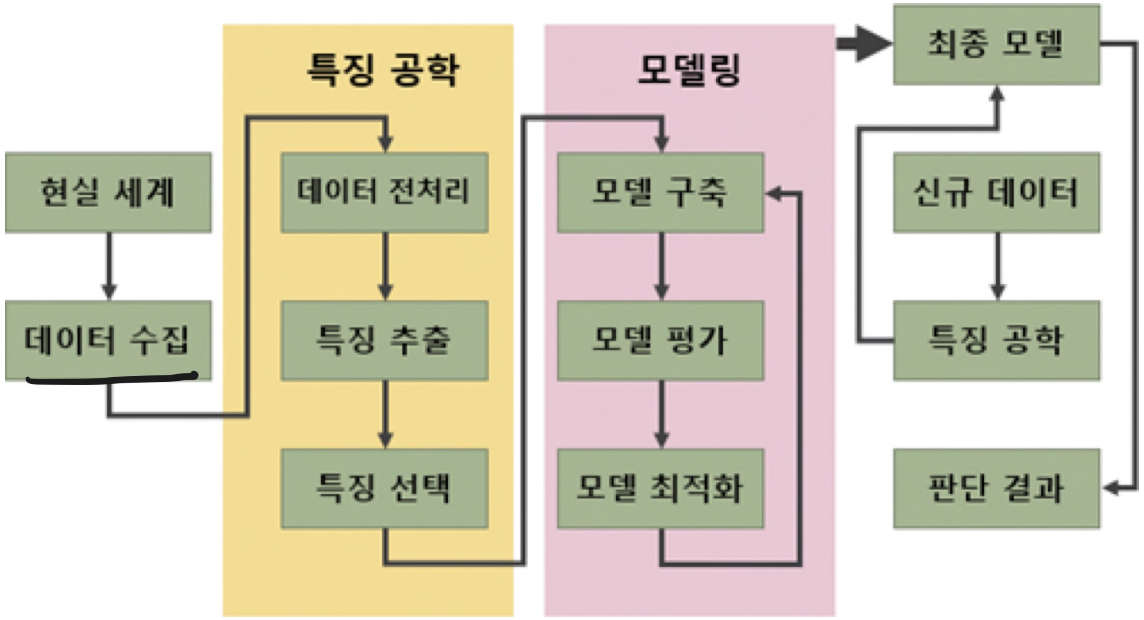
Supervised Learning

1. Model Selection
2. Train Model
3. Evaluate Model

Supervised Learning



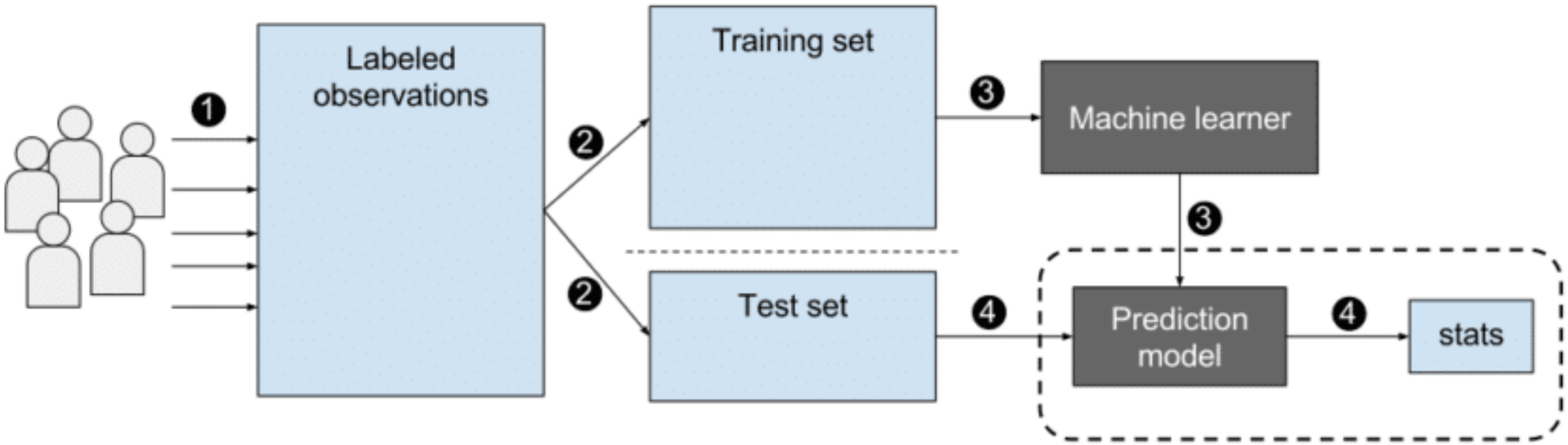
Supervised Learning




```python
from sklearn.linear_model import LogisticRegression

log_reg = LogisticRegression()
log_reg.fit(X, y)
```

Supervised Learning



Supervised Learning

1 큰 그림을 봅니다.
2 데이터를 구합니다.
3 데이터로부터 통찰을 얻기 위해 탐색하고 시각화합니다.
4 머신러닝 알고리즘을 위해 데이터를 준비합니다.
5 모델을 선택하고 훈련시킵니다.
6 모델을 상세하게 조정합니다.
7 솔루션을 제시합니다.
8 시스템을 론칭하고 모니터링하고 유지 보수합니다.

- 모델 선택 → 모델 훈련 → 모델 평가

Loss Function

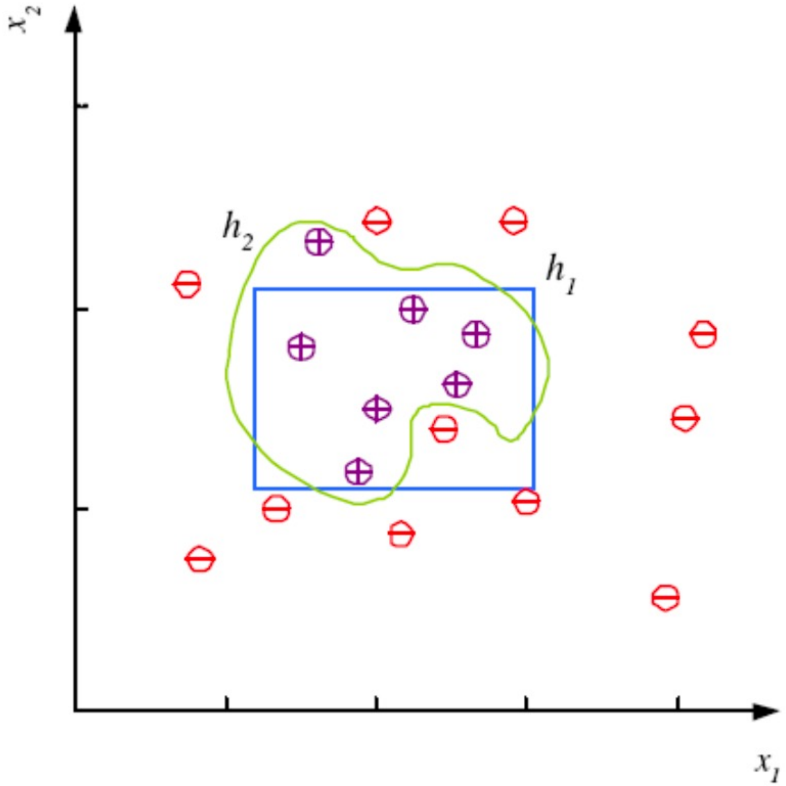
Evaluation Metrics

1. Select Model

Model Selection

Class C of a “family car”

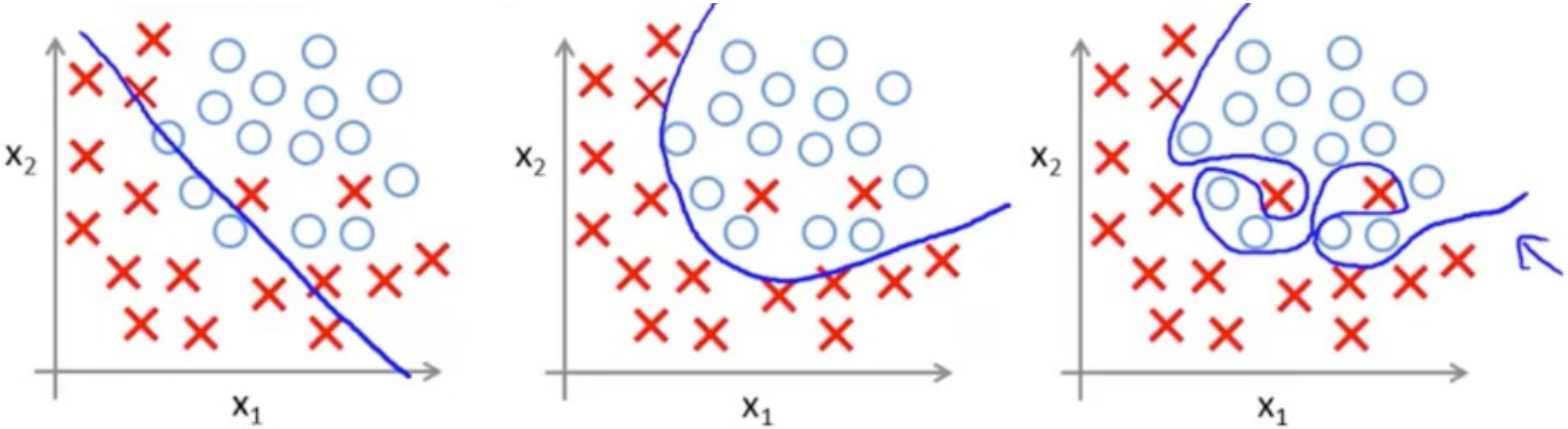
- Input : x1 : price / x2 : engine power
- Output: + or -



→ Assumptions about model is needed!

Model Selection & Generalization

- Underfitting VS Overfitting



2. Train Model

Dimensions of a Supervised Learner

- 모델을 훈련시킨다 = 모델이 훈련 세트에 가장 잘 맞도록 모델 파라미터를 설정

1. Model: $g(\mathbf{x}|\theta)$

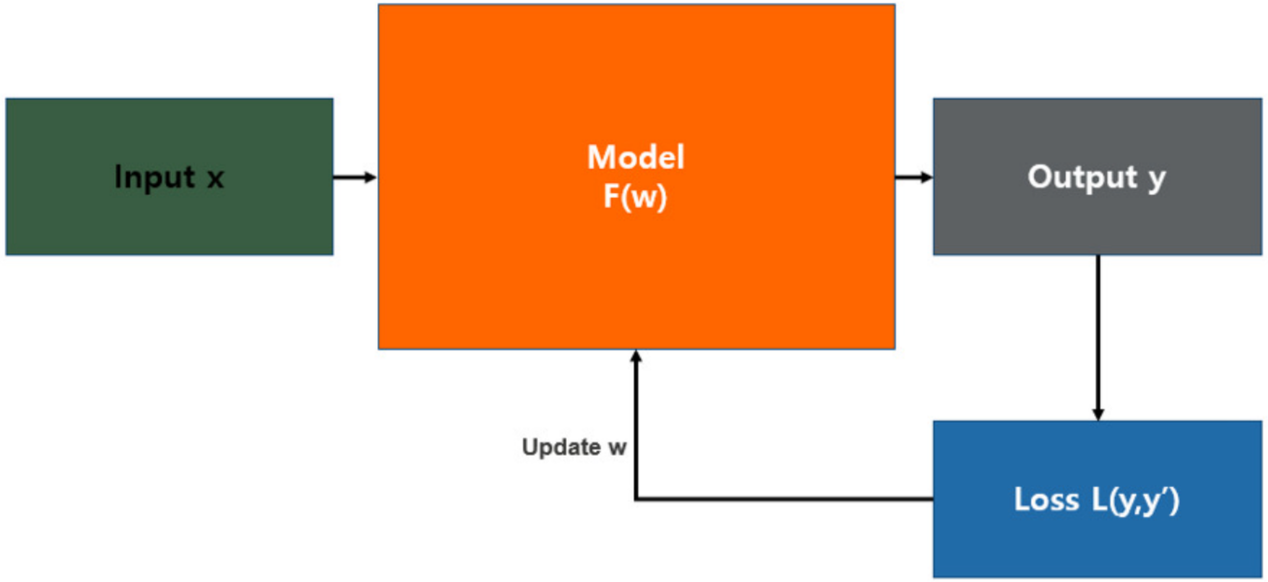
2. Loss function: $E(\theta|\mathcal{X})=\sum_t L\left(r^t, g\left(\mathbf{x}^t|\theta\right)\right)$

3. Optimization procedure: $\theta^* = \arg\min_\theta E(\theta|\mathcal{X})$

Loss Function

- What is Loss Function? 예측값과 실제값(레이블)의 차이를 구하는 기준

Quantifies the error between output of the algorithm and given target value.



Common Loss Functions in ML

Loss function penalizes bad predictions.

Regression

- Mean Squared Error

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

Others:
Mean absolute error and mean bias error

Classification

- Binary Cross Entropy

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log(\hat{y_i}) + (1 - y_i) \cdot \log(1 - \hat{y_i})$$

- Categorical Cross Entropy

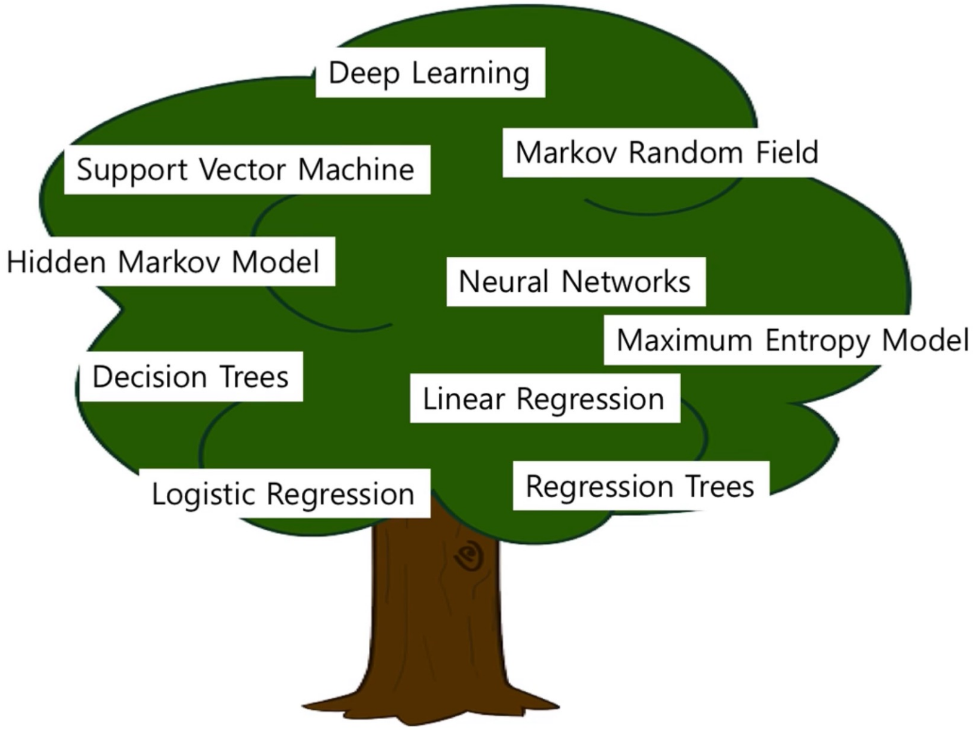
$$CCE = -\frac{1}{N}\sum_{i=0}^{N}\sum_{j=0}^{J} y_j \cdot \log(\hat{y_j}) + (1 - y_j) \cdot \log(1 - \hat{y_j})$$

Others:
Hinge loss / SVM loss.

https://gombru.github.io/2018/05/23/cross_entropy_loss/

Maximum Likelihood Estimator

Maximum Likelihood Estimation



최대 우도 추정 (Maximum Likelihood Estimation)

Likelihood Function

- What is Likelihood Function?

Definition (Likelihood)

For $X_1, \cdots, X_n \overset{iid}{\sim} f_X(x;\theta)$, where θ denotes a parameter of interest. The likelihood function is

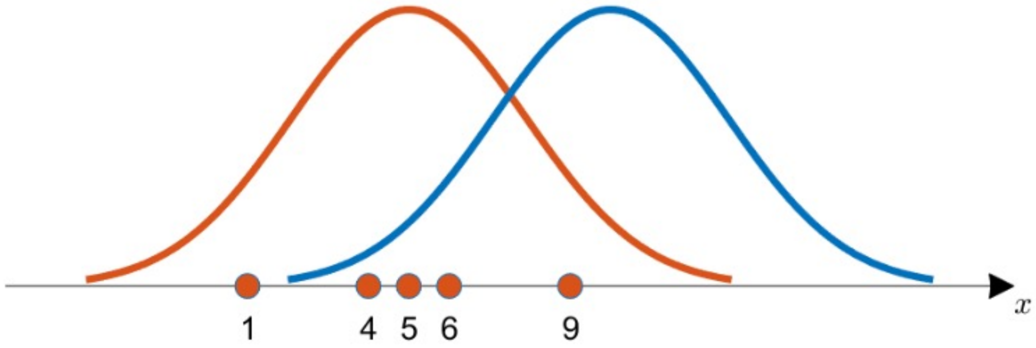
$$L(\theta;\mathbf{X}) = L(\theta; X_1, \cdots, X_n) = \prod_{i=1}^{n} f_X(X_i;\theta)$$

Likelihood Function

다음과 같이 5개의 데이터를 얻었다고 가정하자.

$$x = \{1, 4, 5, 6, 9\}$$

이 때, 아래의 그림을 봤을 때 데이터 x는 주황색 곡선과 파란색 곡선 중 어떤 곡선으로부터 추출되었을 확률이 더 높을까?



Log Likelihood Function

- **Bernoulli distribution**

$$\log L(p) = \sum_{i=1}^{n} (y_i \log p + (1 - y_i) \log (1 - p))$$

- **Binomial distribution**

$$\log L(p) = \log \binom{n}{c} + \sum_{i=1}^{n} (y_i \log p + (1 - y_i) \log (1 - p))$$

- **Multinomial distribution**

$$\log L(p) = \sum_{i=1}^{n} \sum_{j=1}^{c} y_{ij} \log p_j$$

- **Normal distribution**

$$\log L(\mu) \approx - \frac{\sum_{i=1}^{n} (y_i - \mu)}{\sigma^2}$$

Maximum Likelihood Estimator

- What is MLE?

Definition (Maximum likelihood estimator, MLE)

For $X_1, \cdots, X_n \overset{iid}{\sim} f_X(x;\theta)$, the MLE of θ is

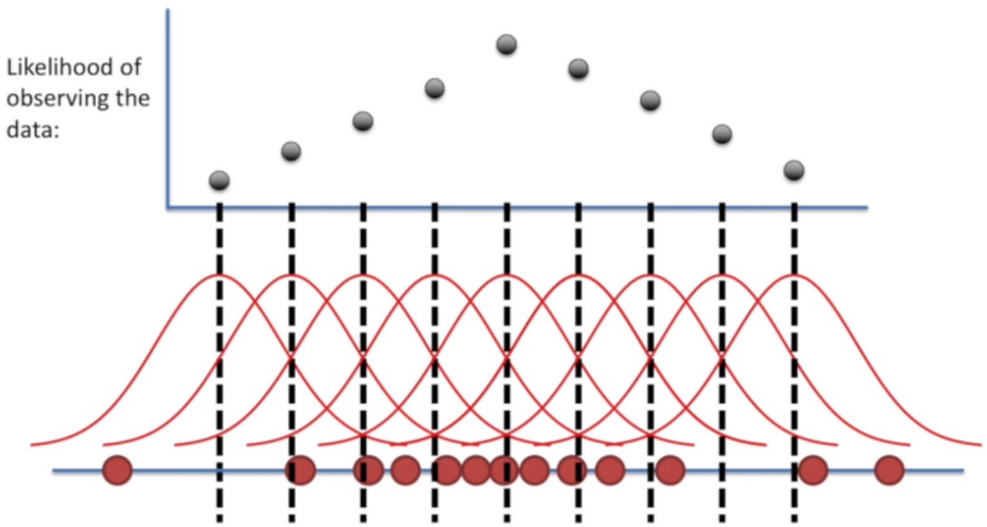
$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, L(\theta; \mathbf{x}).$$

which is equivalent to maximize the logarithm of $L(\theta; \mathbf{x})$ which we call the log-likelihood

$$\ell(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x}).$$

Maximum Likelihood Estimator

- What is MLE?



Binary Classification

- Logistic Regression

Why not Linear Regression?

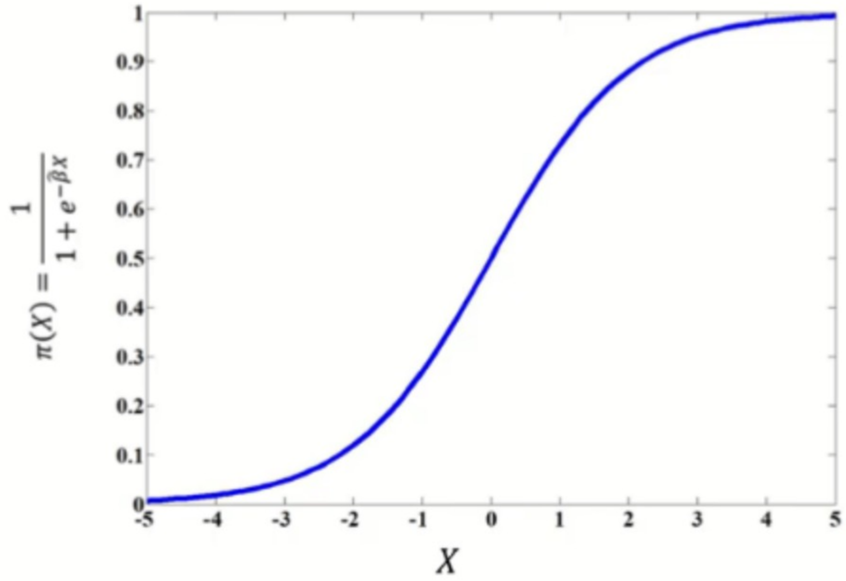
선형 회귀를 통한 분류

- $$Y = \begin{cases} 1 \text{ if korean;} \\ 2 \text{ if american;} \\ 3 \text{ if japanese;} \end{cases} \qquad Y = \begin{cases} 1 \text{ if american;} \\ 2 \text{ if korean;} \\ 3 \text{ if japanese;} \end{cases}$$
- 일반적인 회귀에서는 **라벨의 순서(크기)에 따라 결과가 달라짐**
- 다른 손실 함수나 모델이 필요함

Logistic Regression - Model

$$P(y_i = 1) = \pi_i$$
$$P(y_i = 0) = 1 - \pi_i$$

$$E(y) = \pi(X = x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x) \; = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



Logistic Regression

$$f_i(y_i) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}, i = 1,2,\cdots,n$$

$$L = \prod_i f_i(y_i) = \prod_i \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$$

Log likelihood function : $$\sum_{i=1}^{N} y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i)$$

Logistic Regression – Loss Function

- Cross Entropy = - Log Likelihood 의 기댓값

- Binary Cross Entropy

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

- Log Likelihood 을 최대 = Cross Entropy를 최소

3. Model Evaluation

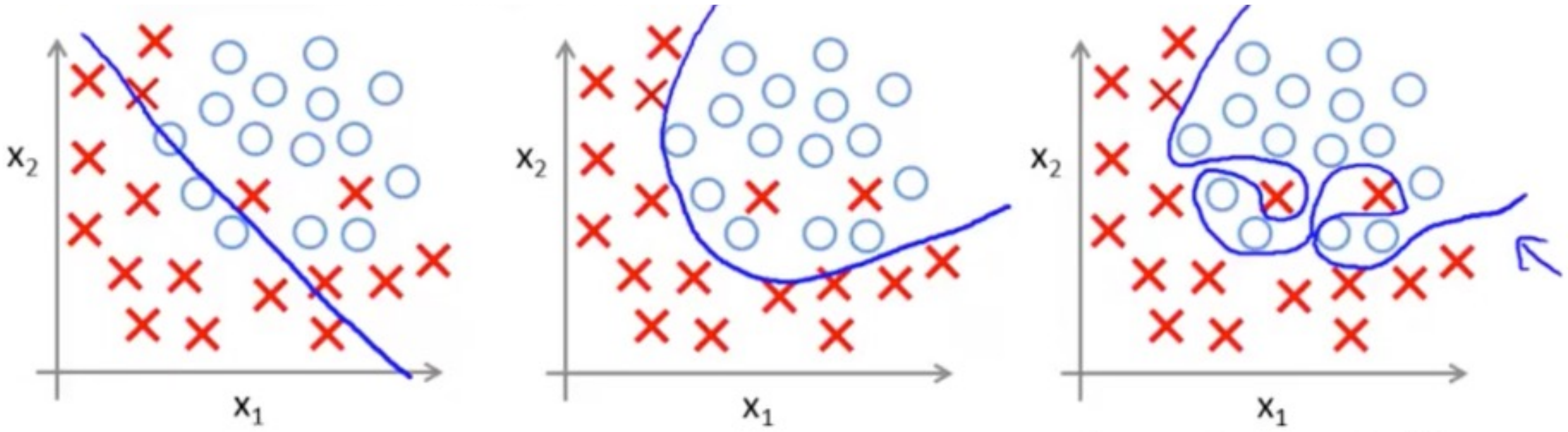
1. 일반화 정확도를 추정

2. 모델의 하이퍼파라미터 튜닝

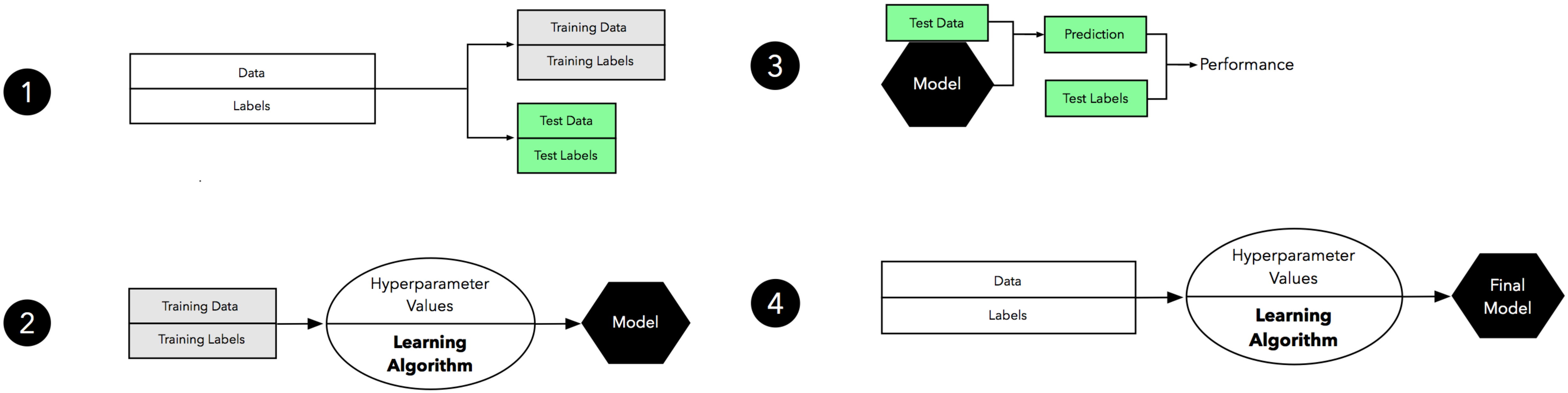
3. 가장 성능이 좋은 알고리즘을 선택

Model Selection & Generalization

- Underfitting VS Overfitting



Hold Out



https://tensorflow.blog/머신-러닝의-모델-평가와-모델-선택-알고리즘-선택-1

3. Model Evaluation

1. 일반화 정확도를 추정
2. **모델의 하이퍼파라미터 튜닝**
3. 가장 성능이 좋은 알고리즘을 선택

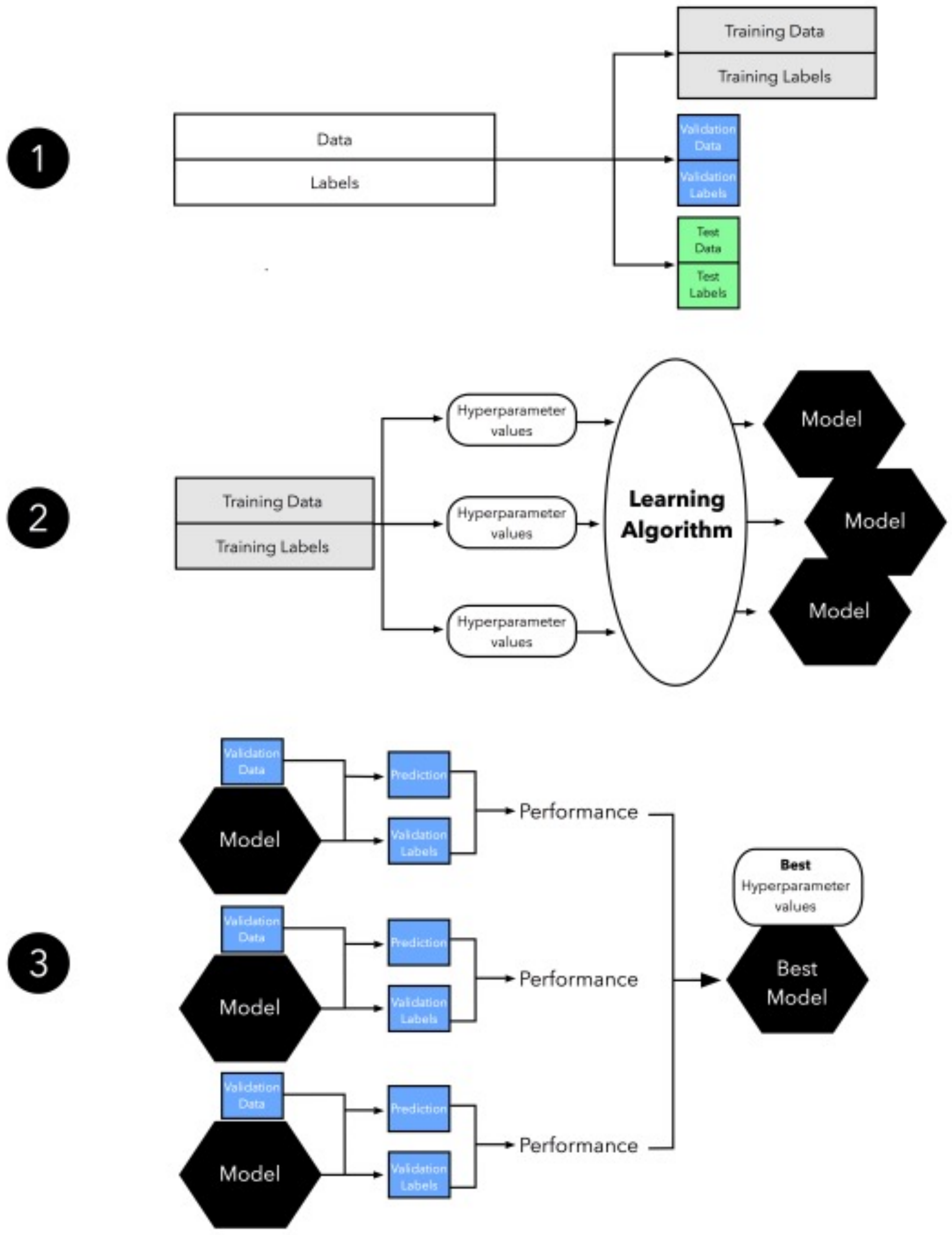
Cross Validation

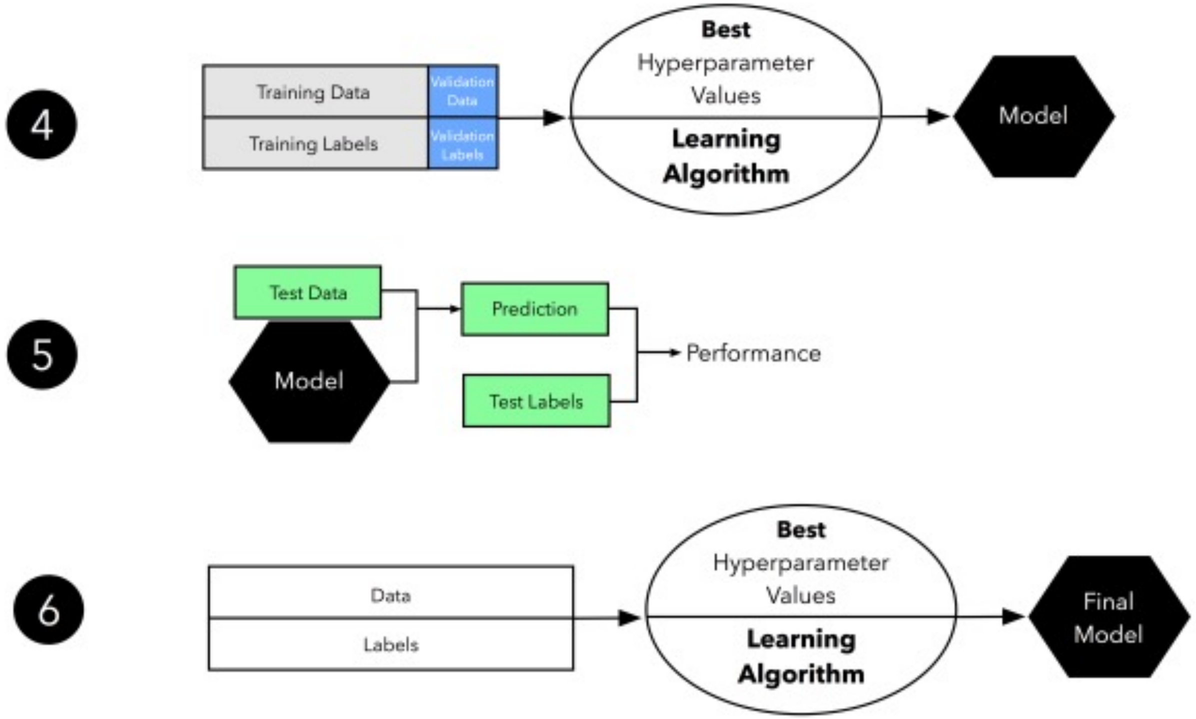
- Cross-Validation

  To estimate generalization error, we need data unseen during training.

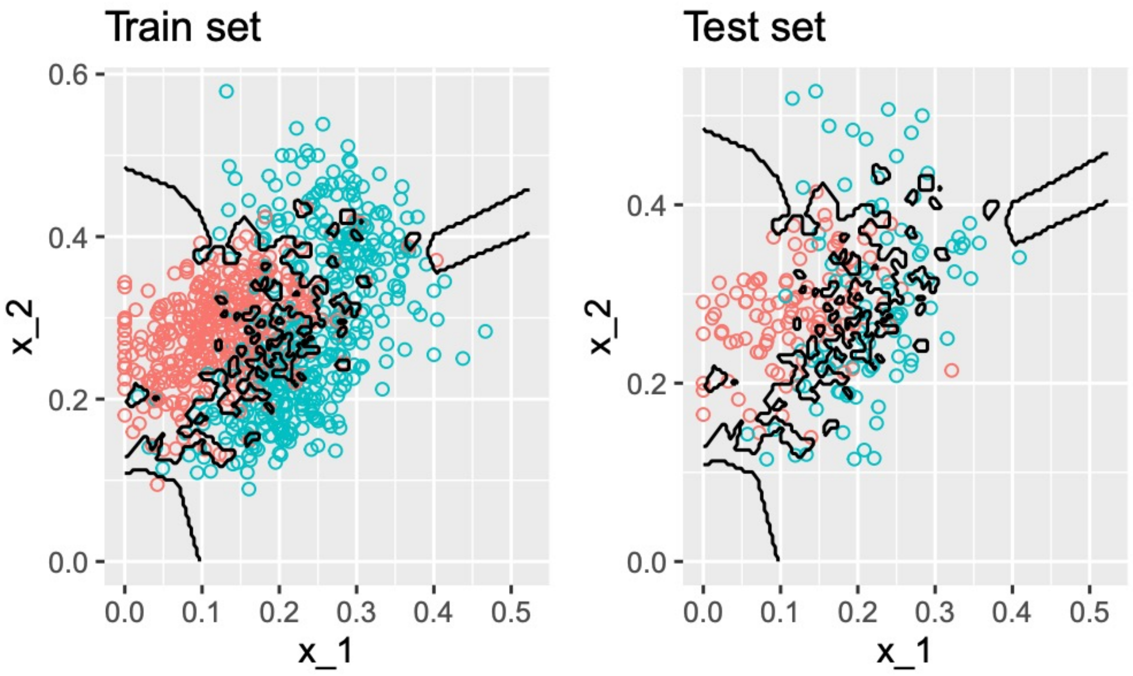
  We split the data as

  - **Training set (50%)**
  - **Validation set (25%)**
  - **Test set (25%)**


1
Data
Labels
Training Data
Training Labels
Validation Data
Validation Labels
Test Data
Test Labels
2
Training Data
Training Labels
Hyperparameter values
Hyperparameter values
Hyperparameter values
Learning Algorithm
Model
Model
Model
3
Validation Data
Model
Prediction
Validation Labels
Performance
Validation Data
Model
Prediction
Validation Labels
Performance
Validation Data
Model
Prediction
Validation Labels
Performance
Best Hyperparameter values
Best Model

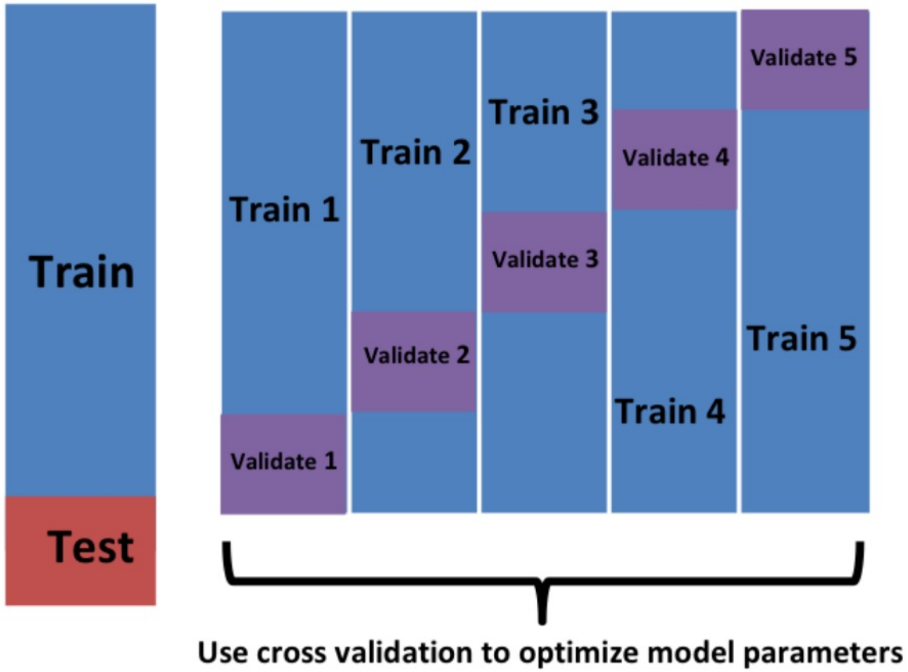

4
Training Data
Training Labels
Validation Data
Validation Labels
Best Hyperparameter Values
Learning Algorithm
Model
5
Test Data
Model
Prediction
Test Labels
Performance
6
Data
Labels
Best Hyperparameter Values
Learning Algorithm
Final Model

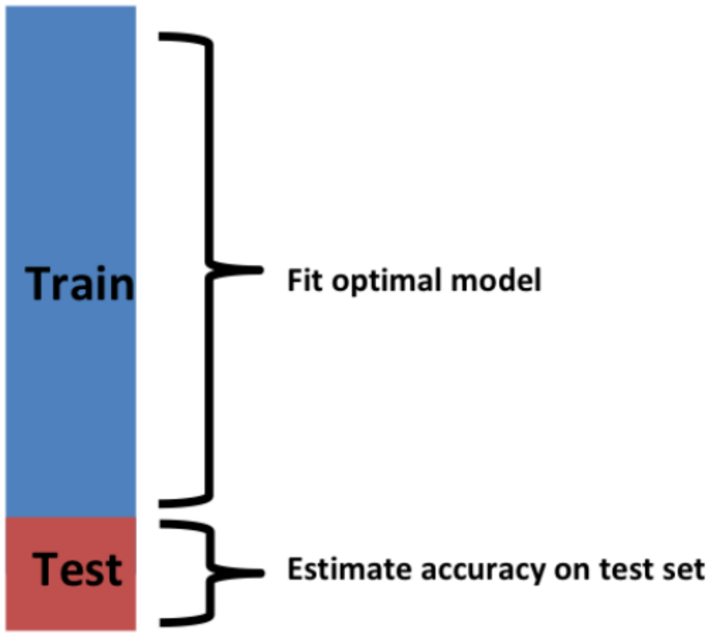
Cross Validation

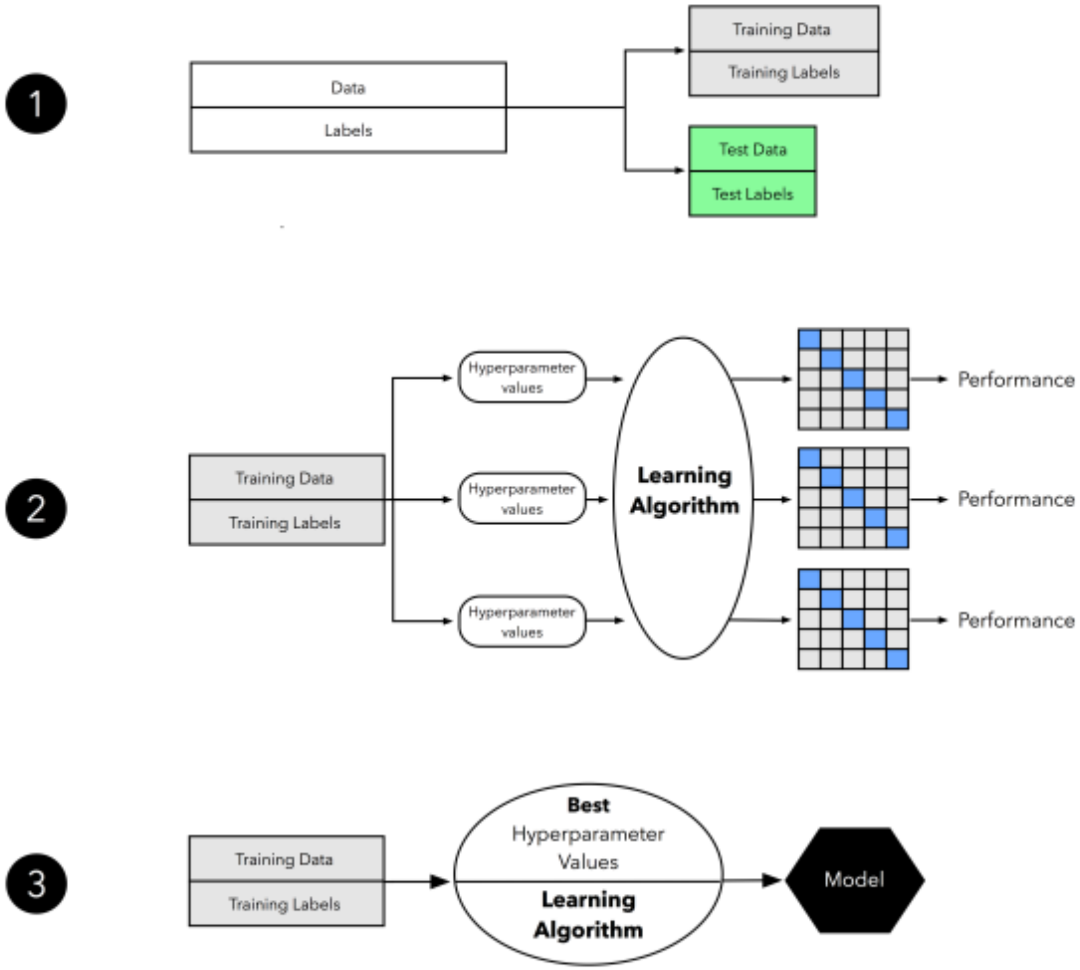


K-Fold Cross Validation

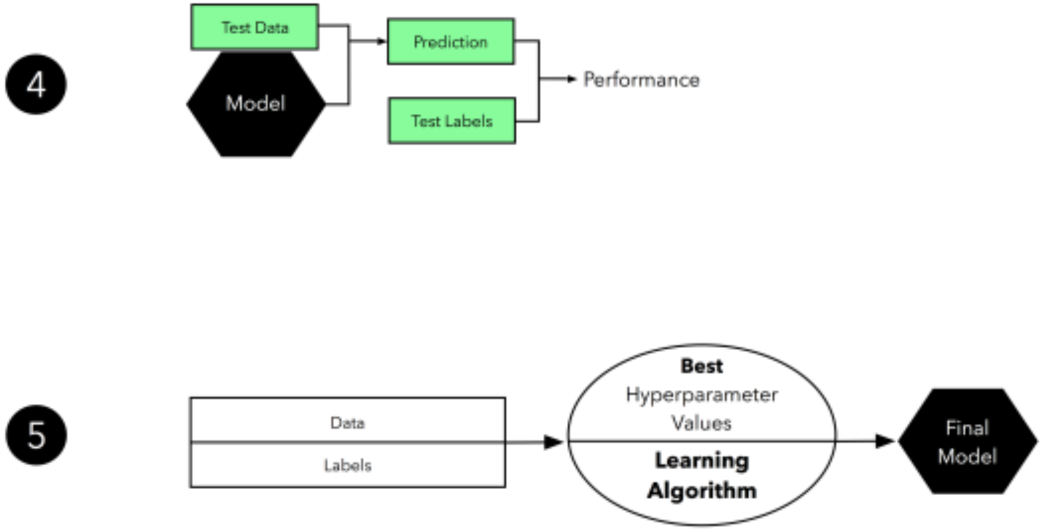
1. Because our data is random, the error is a random variable.
2. If we train an algorithm on the same dataset that we use to compute the apparent error, we might be overtraining.







1
Data
Labels
Training Data
Training Labels
Test Data
Test Labels
2
Training Data
Training Labels
Hyperparameter values
Hyperparameter values
Hyperparameter values
Learning Algorithm
Performance
Performance
Performance
3
Training Data
Training Labels
Best Hyperparameter Values
Learning Algorithm
Model



4
Test Data
Model
Prediction
Test Labels
Performance
5
Data
Labels
Best Hyperparameter Values
Learning Algorithm
Final Model