

LECTURE 7-2

# APPLICATION & TIPS : LEARNING & TEST DATA SETS

Sung Kim <hunkim+ml@gmail.com>  
<http://hunkim.github.io/ml>

**NAVER** | Clova



# **Performance Evaluation : Is This Good?**

# Evaluation Using Training Set?

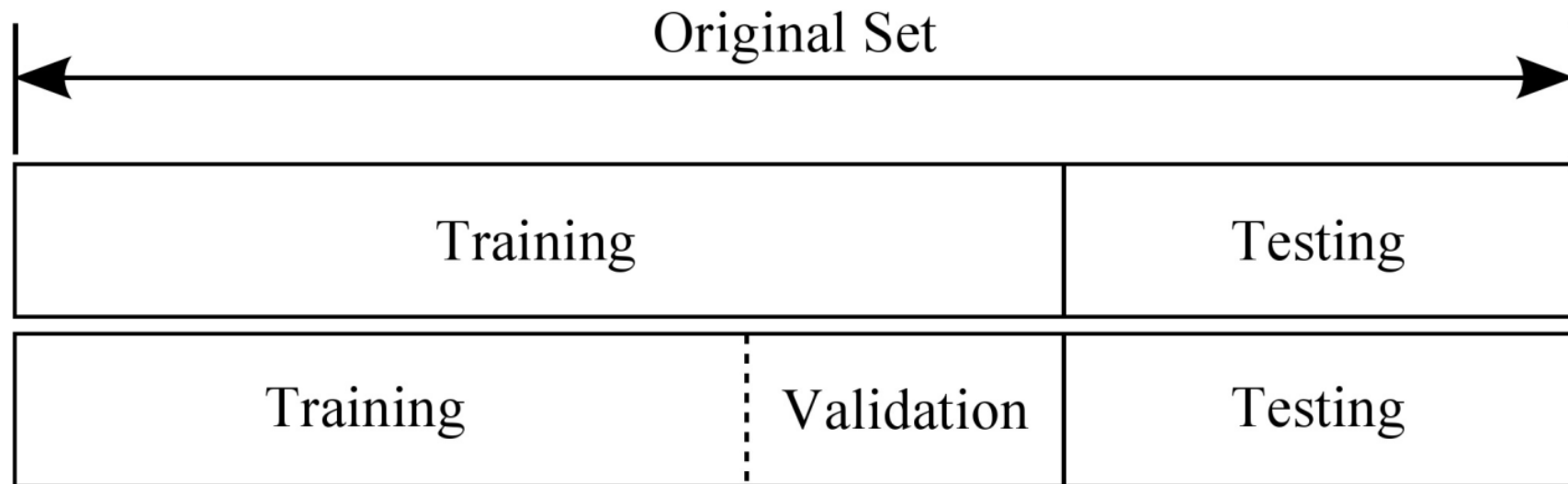
- 100% Correct (Accuracy)
- Can Memorize

Size	Price
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

# Training and Test Sets

Size	Price
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

# Training, Validation and Test Sets



# Online Learning



MODEL

# MNIST Dataset



[train-images-idx3-ubyte.gz](#)

training set images (9912422 bytes)

[train-labels-idx1-ubyte.gz](#)

training set labels (28881 bytes)

[t10k-images-idx3-ubyte.gz](#)

test set images (1648877 bytes)

[t10k-labels-idx1-ubyte.gz](#)

test set labels (4542 bytes)

# Accuracy

- How many of your predictions are correct?
- 95% ~ 99%?
- Check out the lab video



**NEXT LECTURE**

# **DEEP NEURAL NETS!**