SC201 Lecture 5

Training, Validation, Testing Data

- From _____
- Have _____

Big Data

90%	60%	60%	70%
5%	20%	40%	30%
5%	20%		

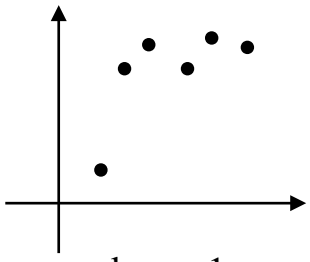
- _____ on Training Data (Dtrain) → Learn _____
- _____ on Validation Data (Dval) (Development) (Ddev) — Choose _____ / _____
- _____ on Testing Data (Dtest) — Tune _____

→ Get _____

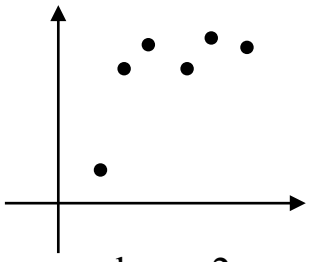
The Problem of Overfitting

- High degree polynomial learns data features _____ (not generalized features)

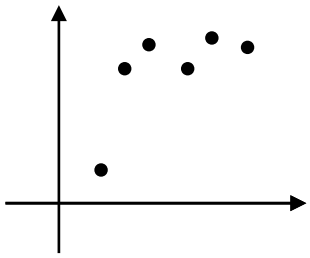
< Linear regression >



degree1



degree2

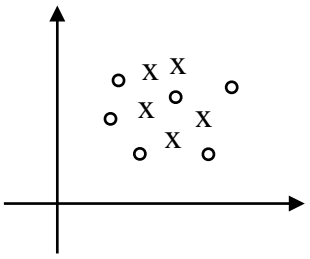


degree n

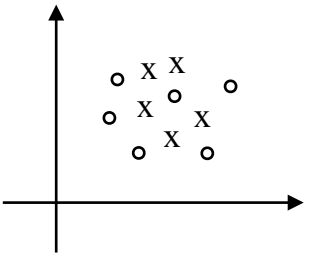
_____ _____ _____

(_____) (_____)

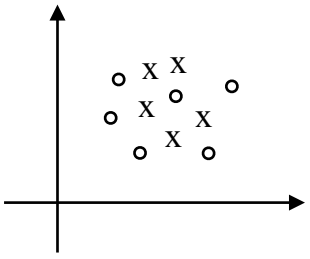
< Logistic regression >



degree1



degree2



degree n

_____ _____

_____ 愈 _____ ➡ _____

_____ 愈 _____ ➡ _____

Address Overfitting

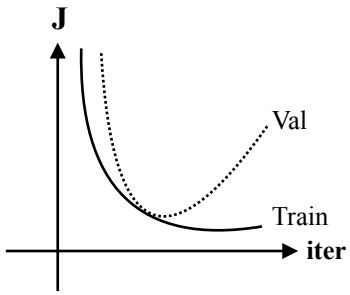
① _____

② _____

③ _____

④ _____

< Early Stopping>



< Cross Validation>

- Split data into ____ folds

  and { train on _____ folds
  validate on ____ folds

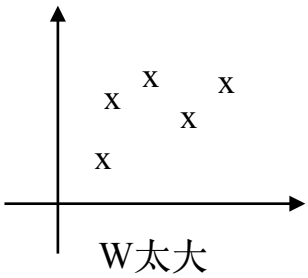
```
for i in range(K):
    Dval = data_folds[i]
    Dtrain = data_folds[:i]+data_folds[i+1: ]
```

< Regularization > (_____)

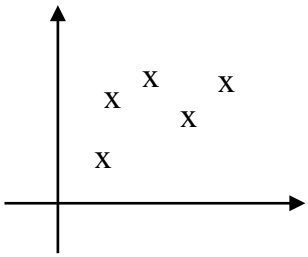
L = [ ] + [ ]

_____ _____ _____

(_____)



W太大



W太小

① W 愈 ____, 彎曲愈 _____

② 但W 愈 ____, _____愈 ____

③ 為了降低Total Loss,

要降低 _____, 也

要降低 _____

④ W太大 ➡ _____

W太小 ➡ _____

$$\frac{dL}{dWi} = \_\_\_\_\_ + \_\_\_\_\_$$

➡ Wi = Wi - _____

< Weight Decay >

Wi = Wi - α((h-y)Xi + λWi)

Wi = Wi - _____

Wi = _____

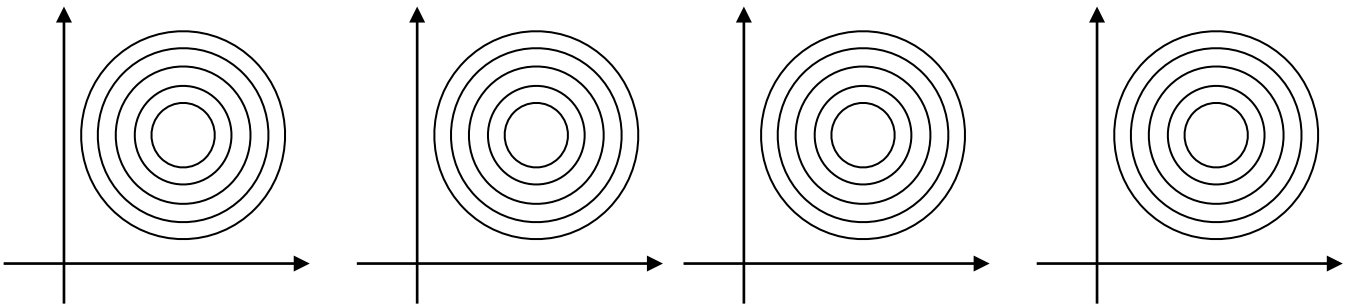
= _____

Bias & Variance

Train Error	1%	15%	15%
Val Error	11%	16%	30%

⬇ ⬇ ⬇

_____ _____ _____



_____variance _____variance _____variance _____variance

_____bias _____bias _____bias _____bias

< fix high bias problem >

- _____ / _____
- _____

< fix high variance problem >

- _____
- _____

Hyperparameters

- _____
- _____
- _____
- _____
- _____