MACHINE LEARNING

Hung-yi Lee

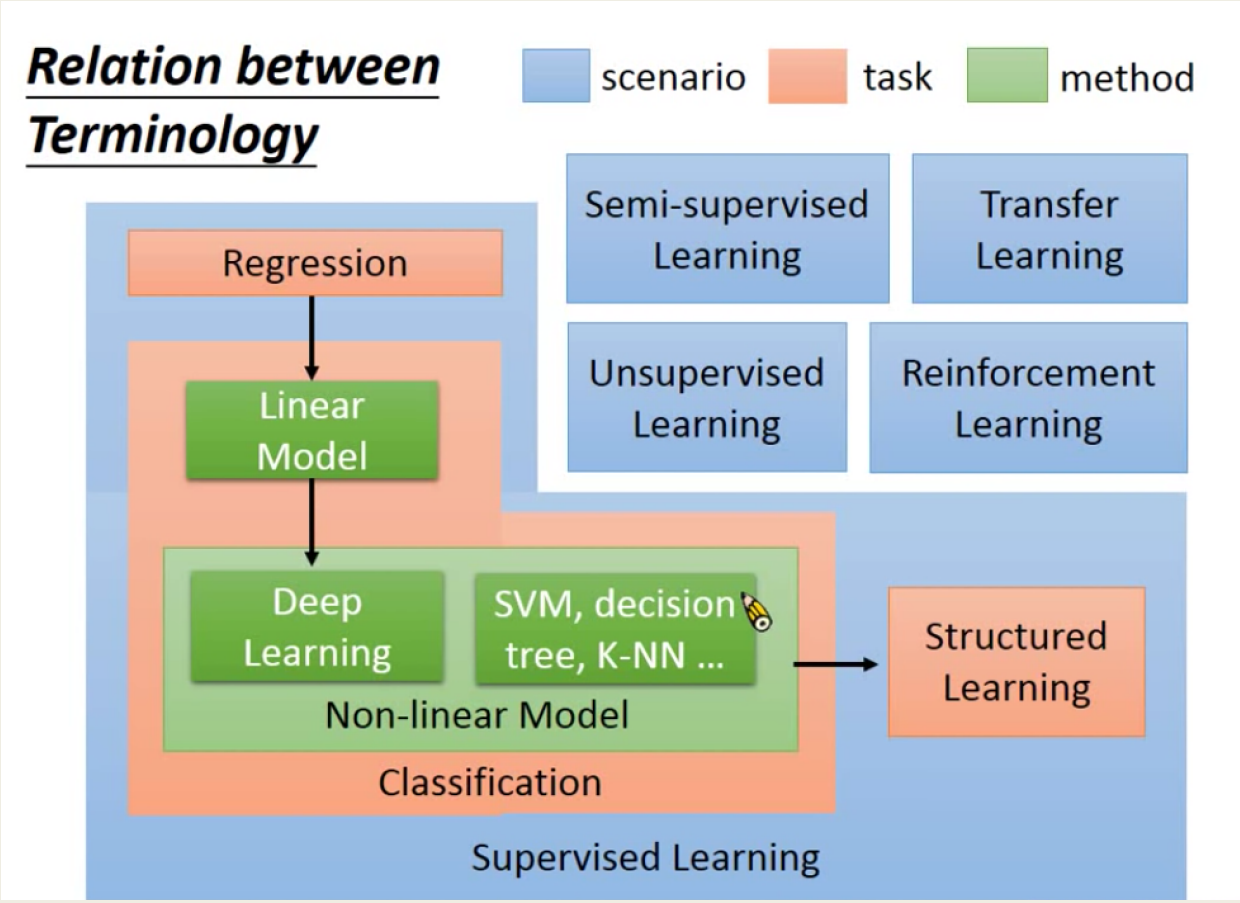
Lecture 0 & 1

分享人：陈晓

目录

- ML 简介
- 回归-Regression
- 复习回顾

ML 简介


Relation between Terminology
scenario
task
method
Semi-supervised Learning
Transfer Learning
Unsupervised Learning
Reinforcement Learning
Regression
Linear Model
Deep Learning
SVM, decision tree, K-NN ...
Non-linear Model
Classification
Structured Learning
Supervised Learning

Regression-回归

- Stock Market Forecast

$f($ $) =$ Dow Jones Industrial Average at tomorrow

- Self-driving Car

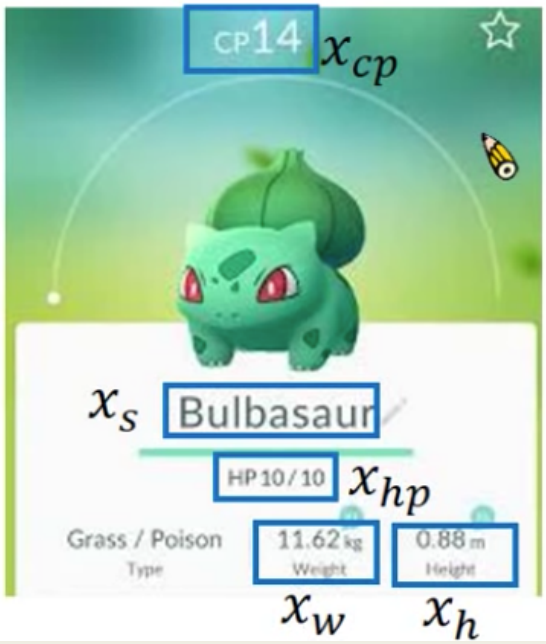
$f($ $) =$ 方向盤角度

- Recommendation

$f($ 使用者 A 商品 B $) =$ 購買可能性

Regression-回归

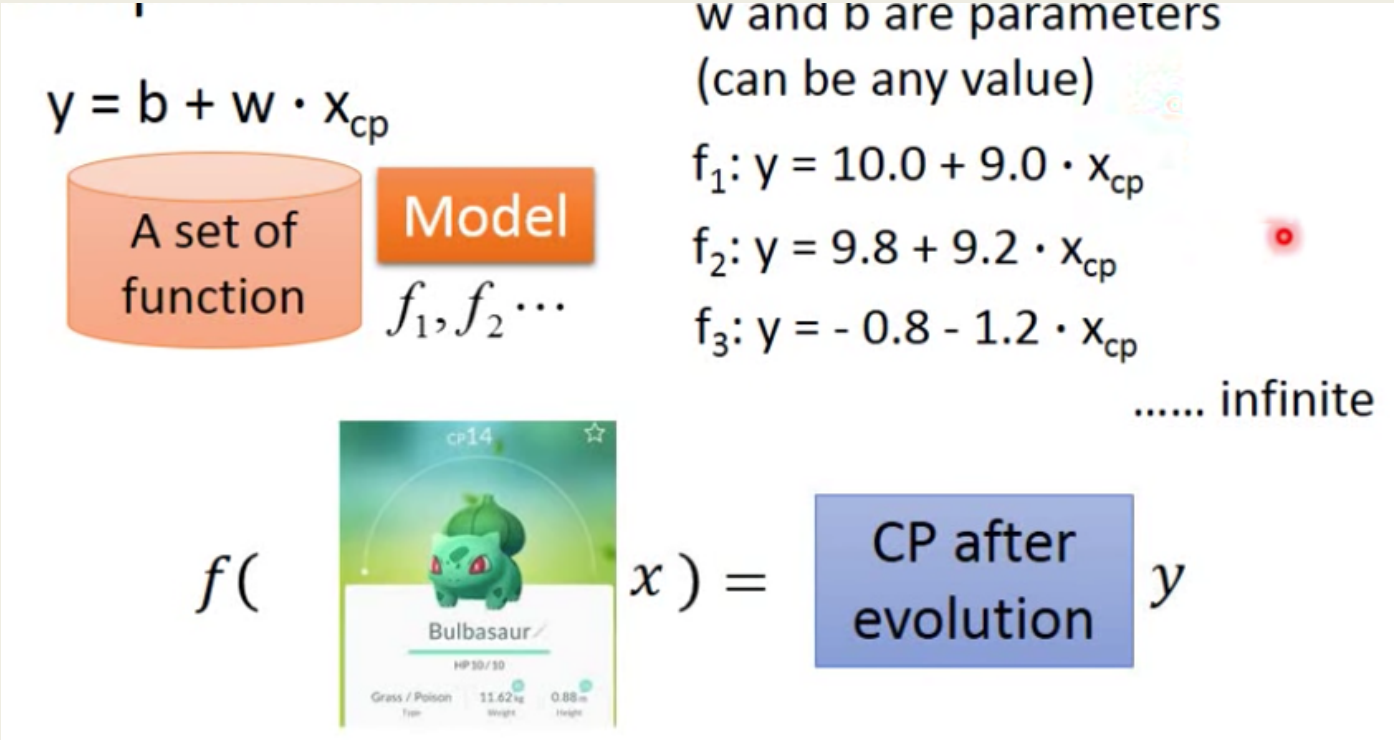
- Estimating the Combat Power (CP) of a pokemon after evolution

$f($ 

x $) =$ CP after evolution y

Regression-回归

1、Model



Regression-回归

2、Data

Training Data:
10 pokemons

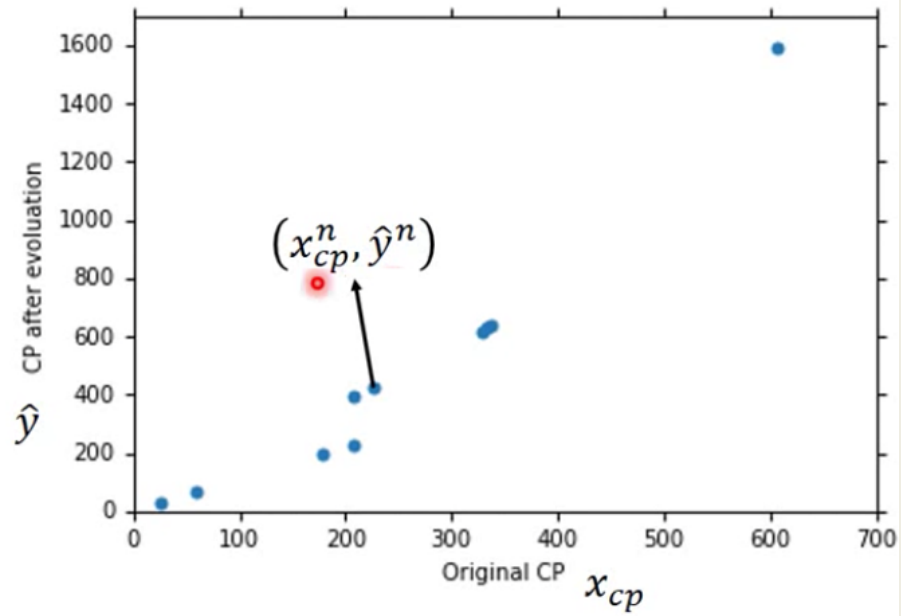
$(x^1, \hat{y}^1)$

$(x^2, \hat{y}^2)$

$\vdots$

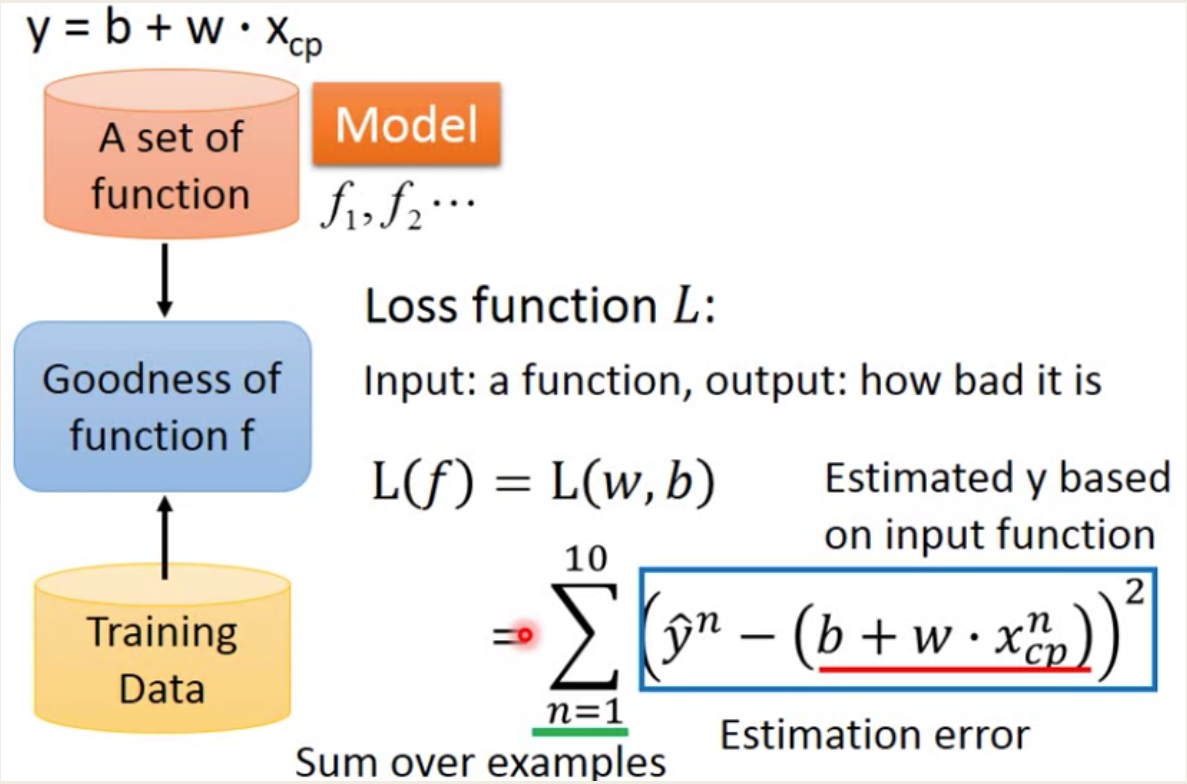
$(x^{10}, \hat{y}^{10})$

This is real data.



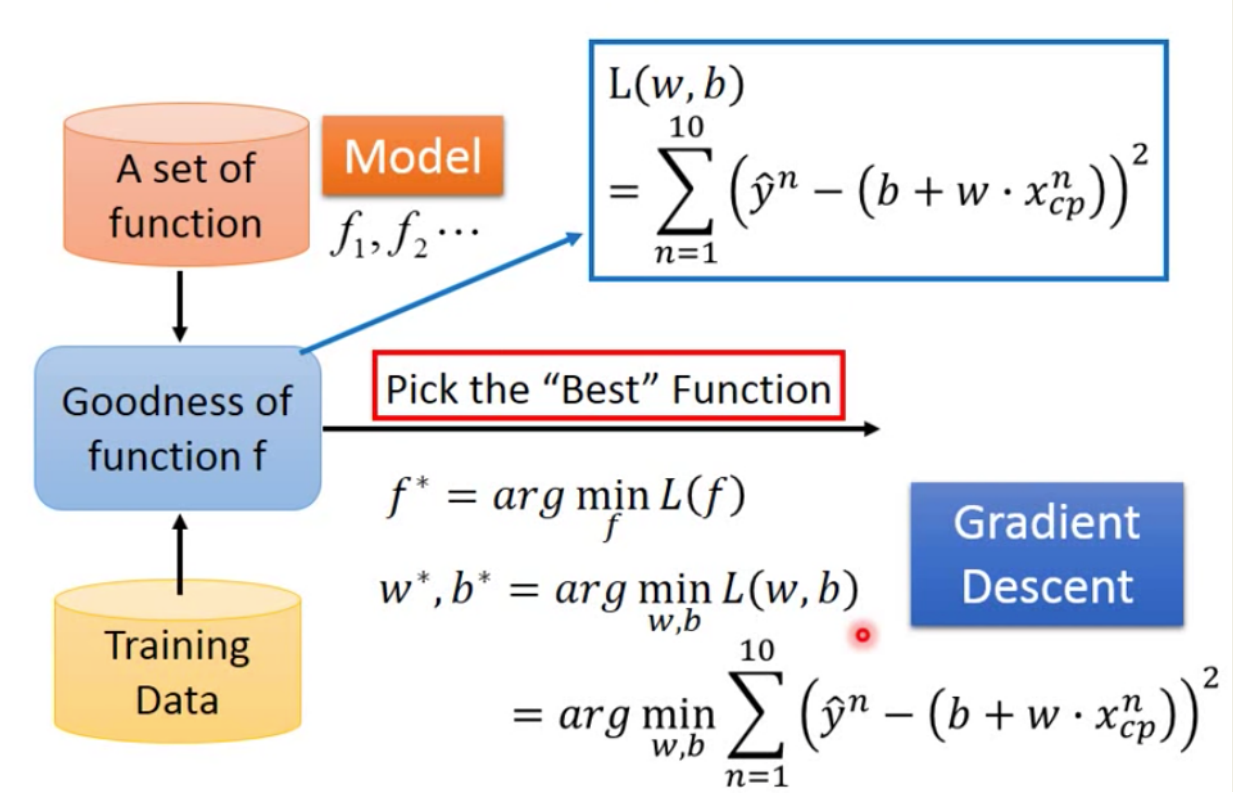
Regression-回归

3、Goodness of function



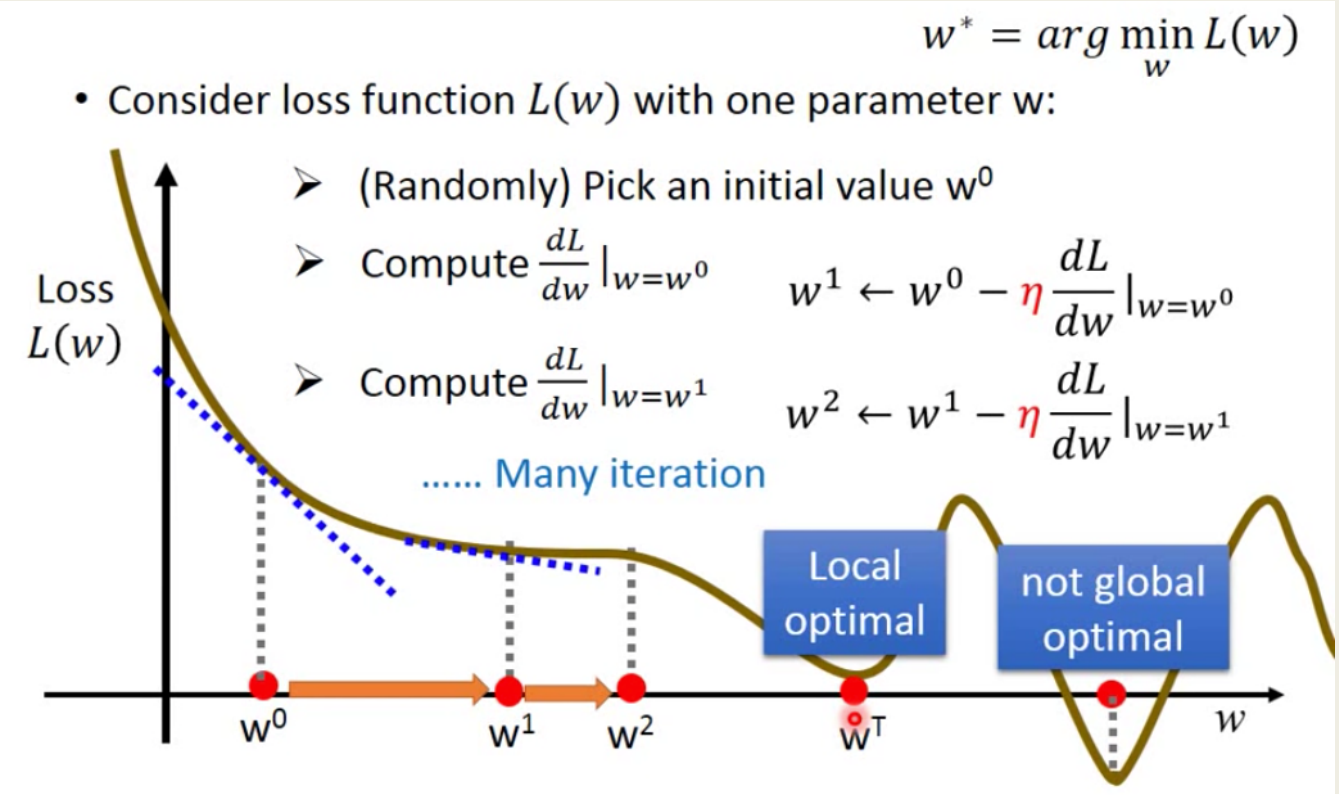
Regression-回归

4、Best function



Regression-回归

5、Gradient Descent（梯度下降）



Regression-回归

5、Gradient Descent（梯度下降）

- How about two parameters? $w^*, b^* = arg \min_{w,b} L(w,b)$
  - (Randomly) Pick an initial value w^0, b^0
  - Compute $\frac{\partial L}{\partial w}|_{w=w^0,b=b^0}$, $\frac{\partial L}{\partial b}|_{w=w^0,b=b^0}$

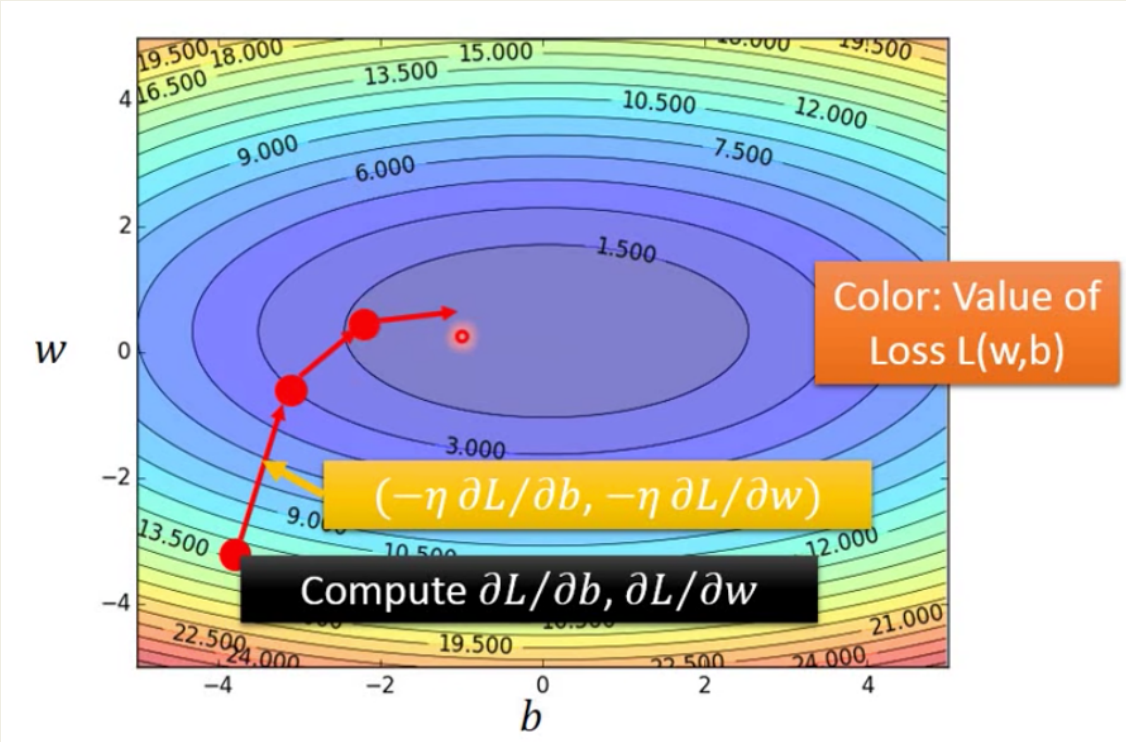
$$w^1 \leftarrow w^0 - \eta \frac{\partial L}{\partial w}|_{w=w^0,b=b^0} \qquad b^1 \leftarrow b^0 - \eta \frac{\partial L}{\partial b}|_{w=w^0,b=b^0}$$

  - Compute $\frac{\partial L}{\partial w}|_{w=w^1,b=b^1}$, $\frac{\partial L}{\partial b}|_{w=w^1,b=b^1}$

$$w^2 \leftarrow w^1 - \eta \frac{\partial L}{\partial w}|_{w=w^0,b=b^0} \qquad b^2 \leftarrow b^1 - \eta \frac{\partial L}{\partial b}|_{w=w^0,b=b^0}$$

Regression-回归

5、Gradient Descent（梯度下降）



Regression-回归

5、Gradient Descent（梯度下降）

- Formulation of $\partial L/\partial w$ and $\partial L/\partial b$

$$L(w,b) = \sum_{n=1}^{10} \left(\hat{y}^n - \left(b + w \cdot x_{cp}^n\right)\right)^2$$

$$\frac{\partial L}{\partial w} =? \sum_{n=1}^{10} 2\left(\hat{y}^n - \left(b + w \cdot x_{cp}^n\right)\right)\left(-x_{cp}^n\right)$$

$$\frac{\partial L}{\partial b} =? \sum_{n=1}^{10} 2\left(\hat{y}^n - \left(b + w \cdot x_{cp}^n\right)\right)(-1)$$

Regression-回归

6、结果——Training

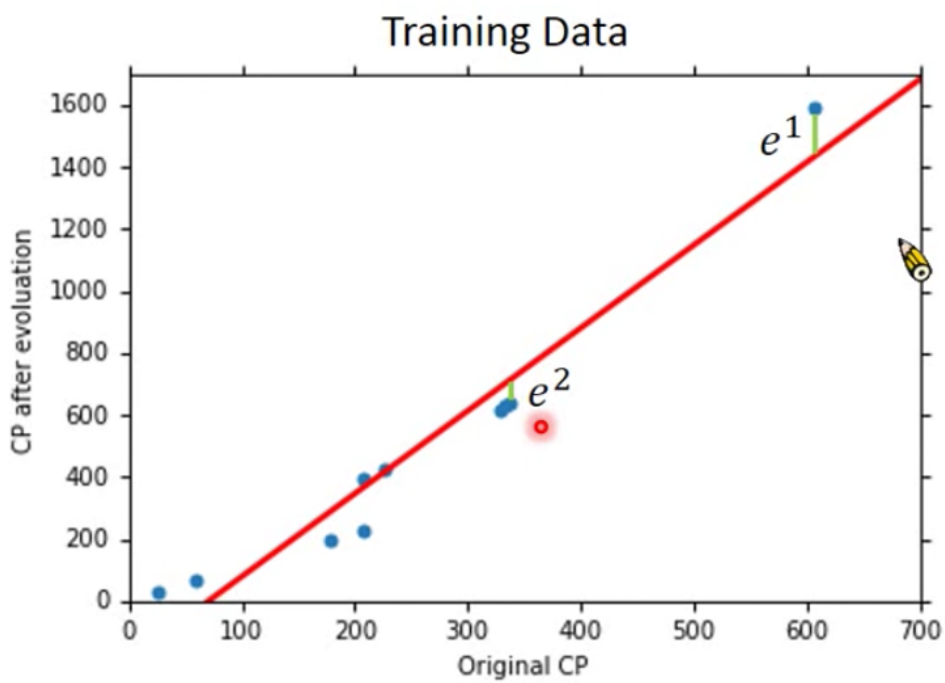
$y = b + w \cdot x_{cp}$

b = -188.4

w = 2.7

Average Error on Training Data

$$= \sum_{n=1}^{10} e^n = 31.9$$



Regression-回归

6、结果——Testing

$y = b + w \cdot x_{cp}$

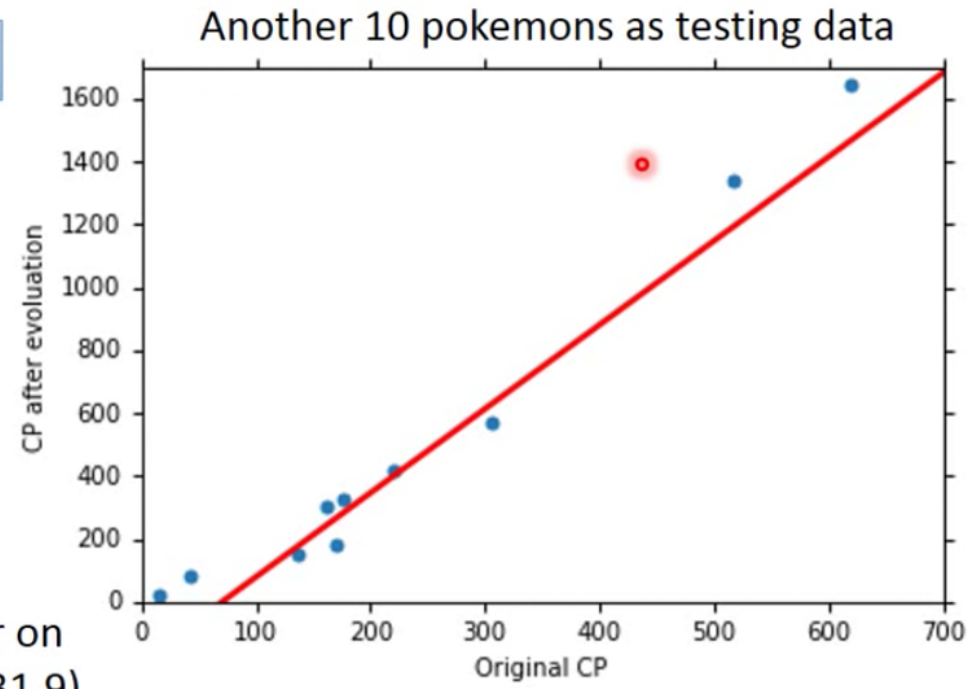
b = -188.4

w = 2.7

Average Error on Testing Data

$$= \sum_{n=1}^{10} e^n = 35.0$$

> Average Error on Training Data (31.9)



Regression-回归

7、改良Model（引入二次式）

Selecting another Model

$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2$$

Best Function

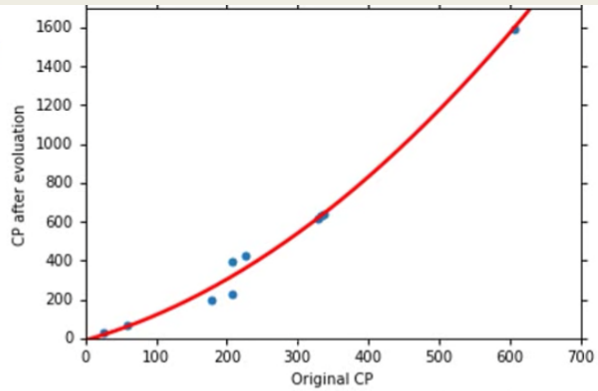
$b = -10.3$

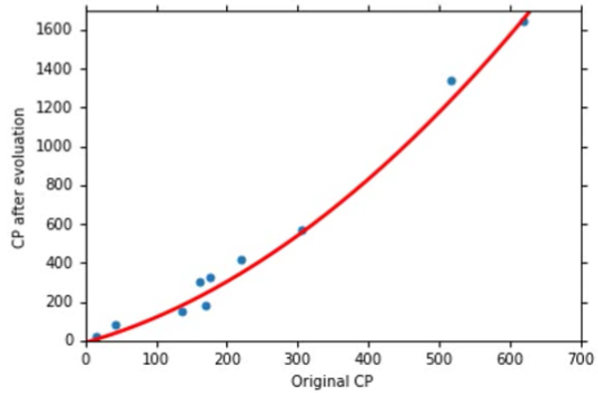
$w_1 = 1.0$, $w_2 = 2.7 \times 10^{-3}$

Average Error = 15.4

Testing:

Average Error = 18.4





Regression-回归

7、改良Model（引入三次式）

Selecting another Model

$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3$$

Best Function

$b = 6.4$, $w_1 = 0.66$

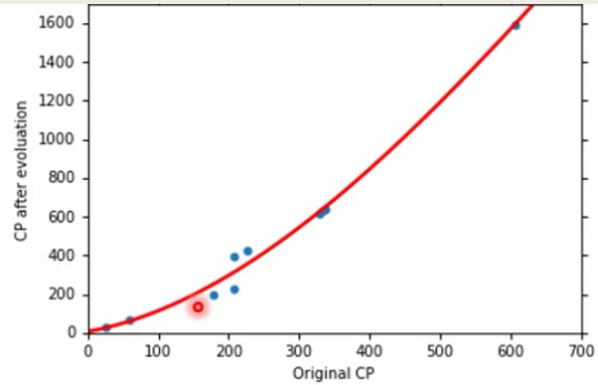
$w_2 = 4.3 \times 10^{-3}$

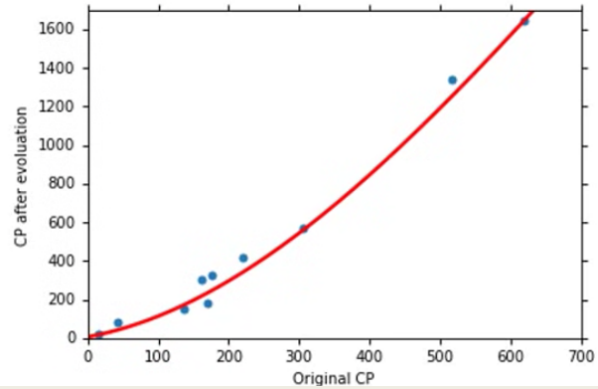
$w_3 = -1.8 \times 10^{-6}$

Average Error = 15.3

Testing:

Average Error = 18.1





Regression-回归

7、改良Model（引入四次式）

Selecting another Model

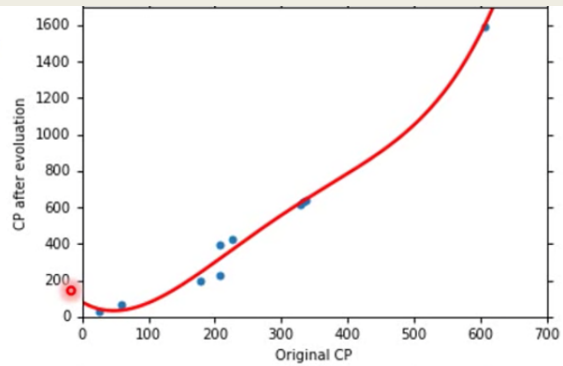
$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4$$

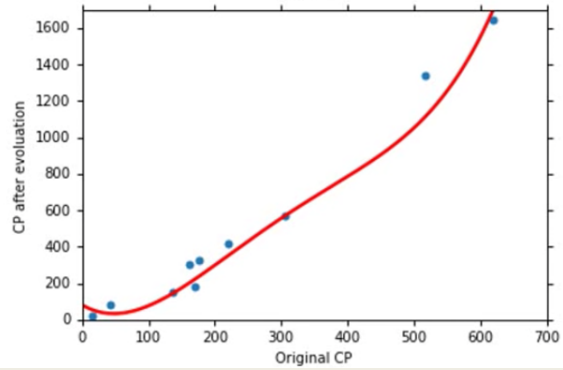
Best Function

Average Error = 14.9

Testing:

Average Error = 28.8



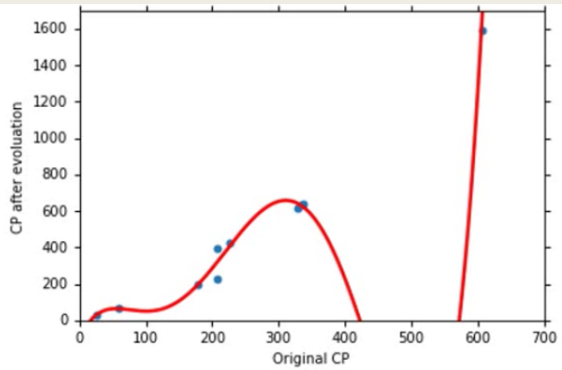


Regression-回归

7、改良Model（引入五次式）

Selecting another Model

$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$$

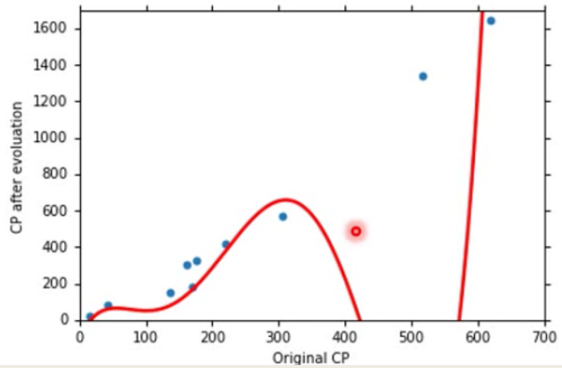


Best Function

Average Error = 12.8

Testing:

Average Error = 232.1



Regression-回归

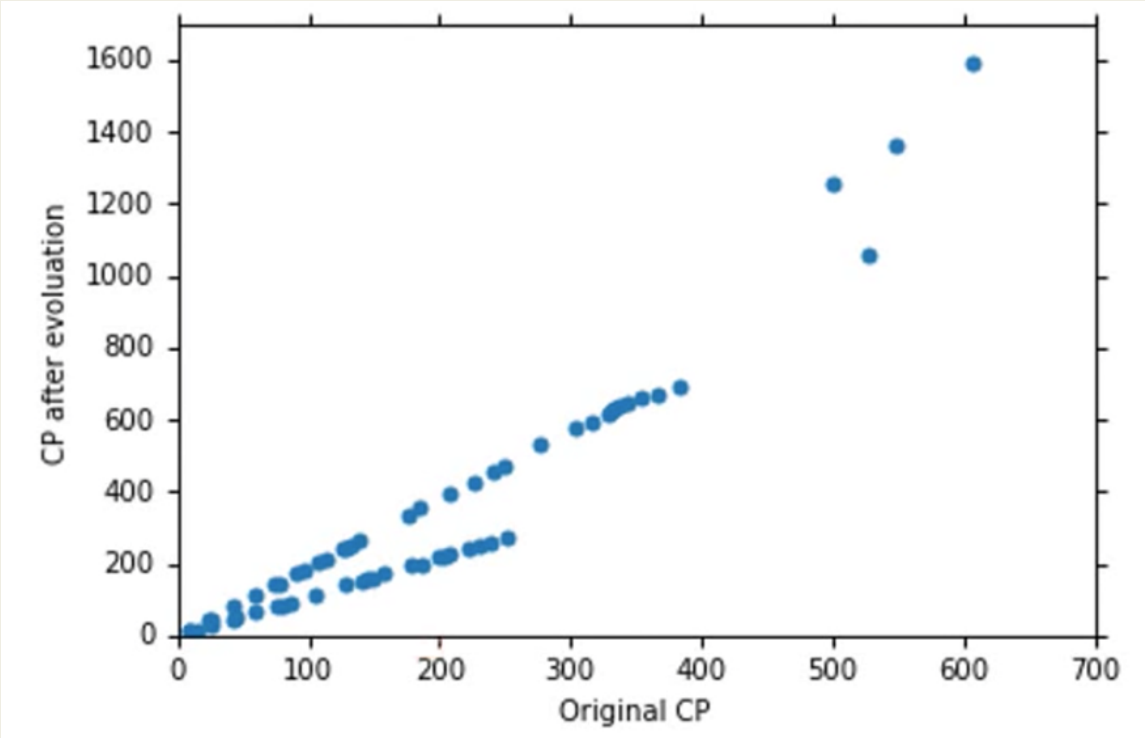
7、改良Model——总结

	Training	Testing
1	31.9	35.0
2	15.4	18.4
3	15.3	18.1
4	14.9	28.2
5	12.8	232.1

A more complex model does not always lead to better performance on ***testing data***.

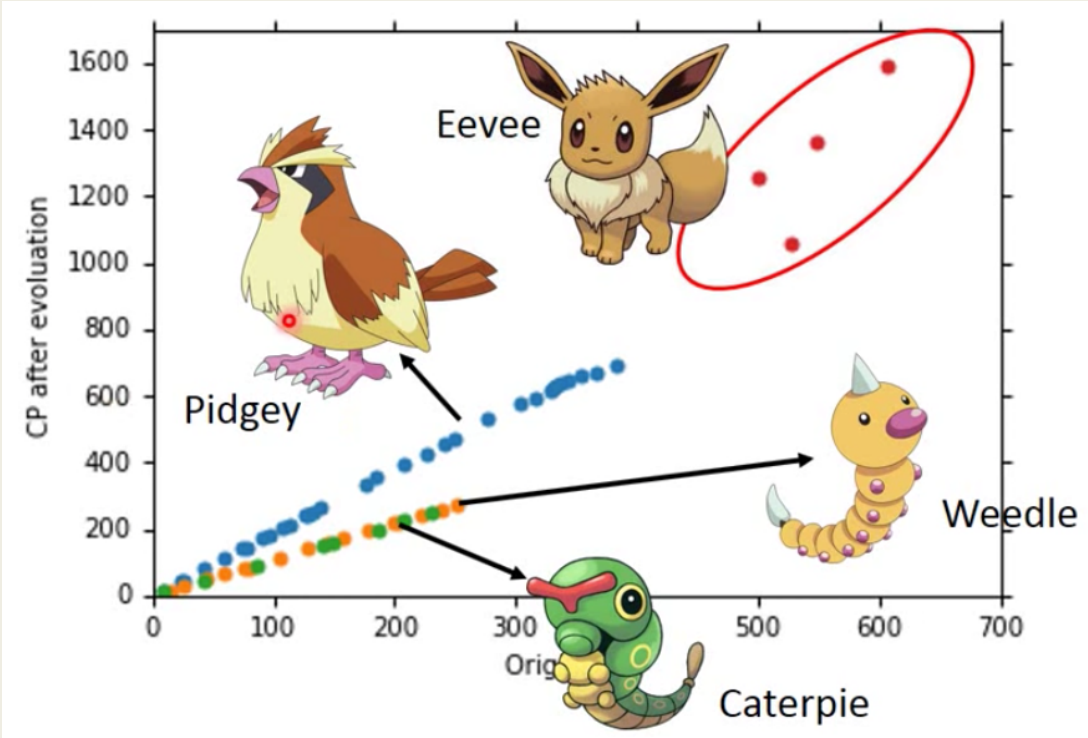
Regression-回归

8、收集更多Data



Regression-回归

8、收集更多Data



Regression-回归

9、重新构造Model

x_s = species of x

x

If x_s = Pidgey: $y = b_1 + w_1 \cdot x_{cp}$

If x_s = Weedle: $y = b_2 + w_2 \cdot x_{cp}$

If x_s = Caterpie: $y = b_3 + w_3 \cdot x_{cp}$

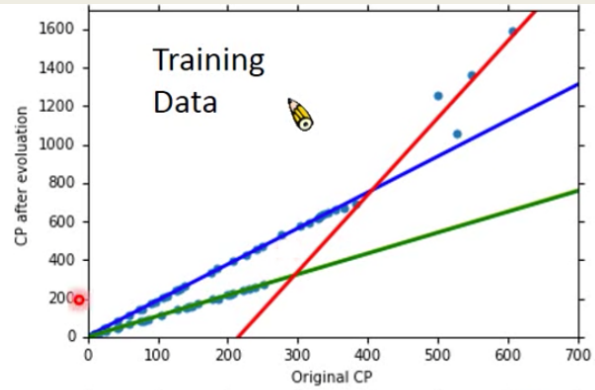
If x_s = Eevee: $y = b_4 + w_4 \cdot x_{cp}$

y

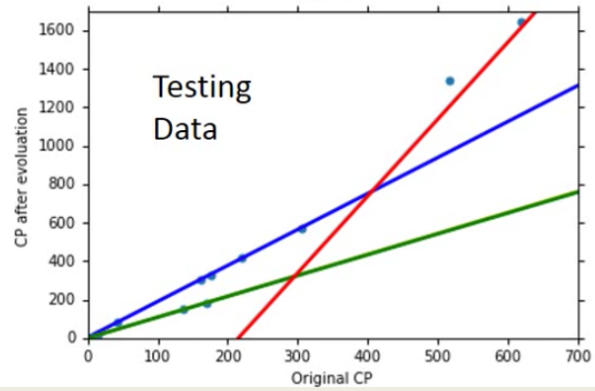
Regression-回归

10、结果

Average error
= 3.8



Average error
= 14.3



Regression-回归

11、重新构造最复杂的Model

x

If x_s = Pidgey: $y' = b_1 + w_1 \cdot x_{cp} + w_5 \cdot (x_{cp})^2$

If x_s = Weedle: $y' = b_2 + w_2 \cdot x_{cp} + w_6 \cdot (x_{cp})^2$

If x_s = Caterpie: $y' = b_3 + w_4 \cdot x_{cp} + w_7 \cdot (x_{cp})^2$

If x_s = Eevee: $y' = b_4 + w_4 \cdot x_{cp} + w_8 \cdot (x_{cp})^2$

$$y = y' + w_9 \cdot x_{hp} + w_{10} \cdot (x_{hp})^2 + w_{11} \cdot x_h + w_{12} \cdot (x_h)^2 + w_{13} \cdot x_w + w_{14} \cdot (x_w)^2$$

y

Training Error = 1.9

Testing Error = 102.3

Overfitting!

Regression-回归

12、Regularization——正则化

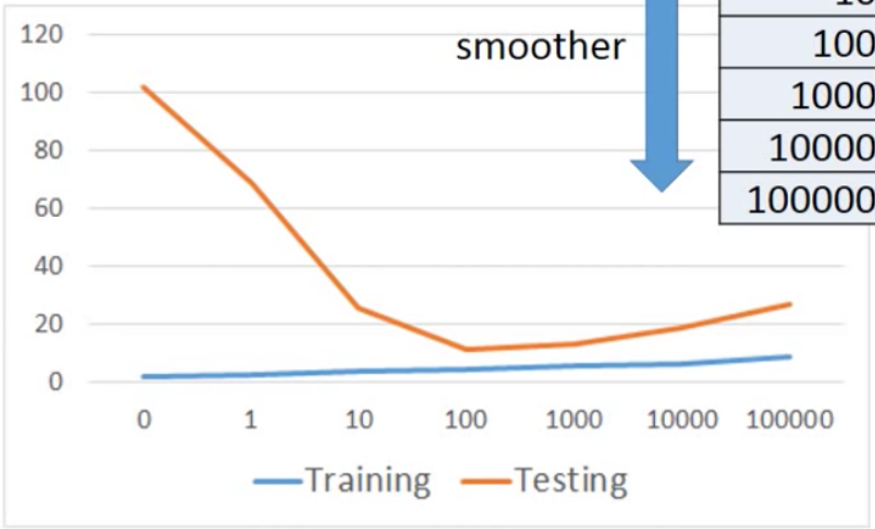
$$y = b + \sum w_i x_i$$

The functions with smaller w_i are better

$$L = \sum_n \left(\hat{y}^n - \left(b + \sum w_i x_i \right) \right)^2 \boxed{+\lambda \sum (w_i)^2}$$

Regression-回归

13、正则化的结果



λ	Training	Testing
0	1.9	102.3
1	2.3	68.7
10	3.5	25.7
100	4.1	11.1
1000	5.6	12.8
10000	6.3	18.7
100000	8.5	26.8

How smooth?

Select λ obtaining the best model

复习回顾

- Lecture 0
  - Supervised Learning
    - Regression
    - Classification
      - Linear Model
      - Non-linear Model
        - Deep Learning
        - SVM、Decision Tree、K-NN.....
    - Structured Learning
  - Semi-supervised Learning
  - Unsupervised Learning
  - Transfer Learning
  - Reinforcement Learning

复习回顾

- **Lecture 1**
  - Regression的定义
  - 选择Model
  - 如何定义Model的好坏
    - Loss Function
  - 梯度下降
  - 改良Model
    - 会导致Overfitting（过拟合）
  - 正则化
    - 考虑更多的变量，更容易导致过拟合

THANK YOU