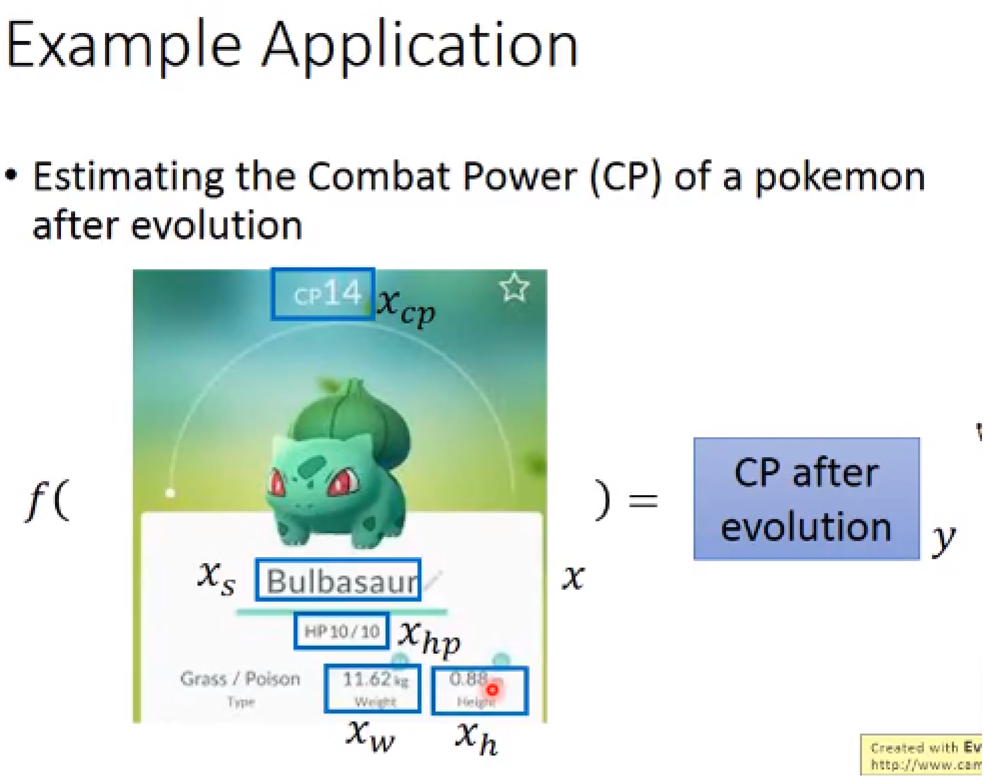
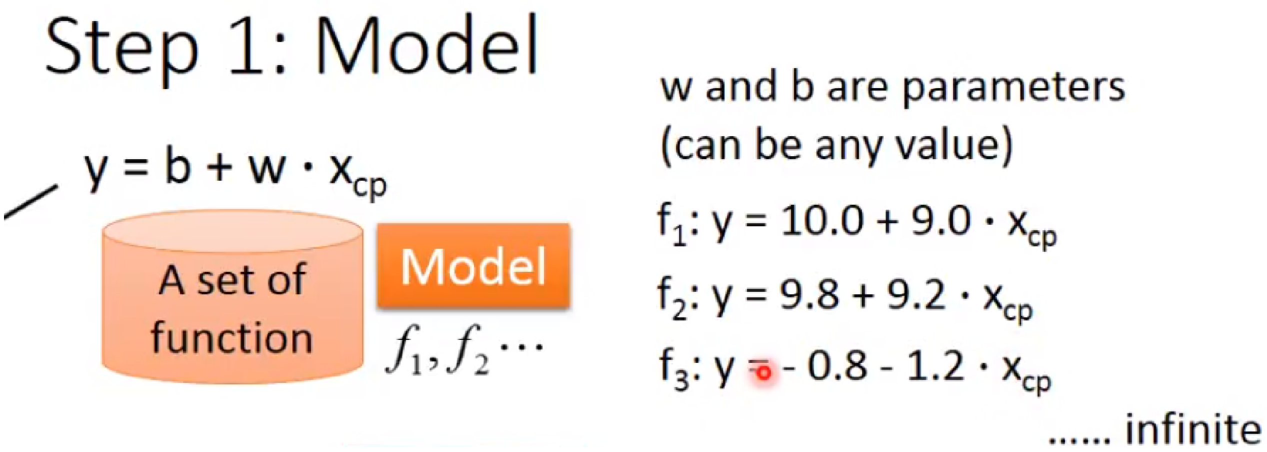
举例


Example Application
• Estimating the Combat Power (CP) of a pokemon after evolution
CP14
x_{cp}
$f($
$) =$
CP after evolution
y
x_s
Bulbasaur
HP 10 / 10
x_{hp}
x
Grass / Poison
Type
11.62 kg
Weight
0.88 m
Height
x_w
x_h

1. 模型确立

假设初始创建模型如下类的表示，则有无穷多的模型公式



而线性模型可以视为

$$y = b + \sum_{i=1}^{n} w_i x_i$$

$$x_i : xcp, xhp, xw, xh \ldots\ldots$$

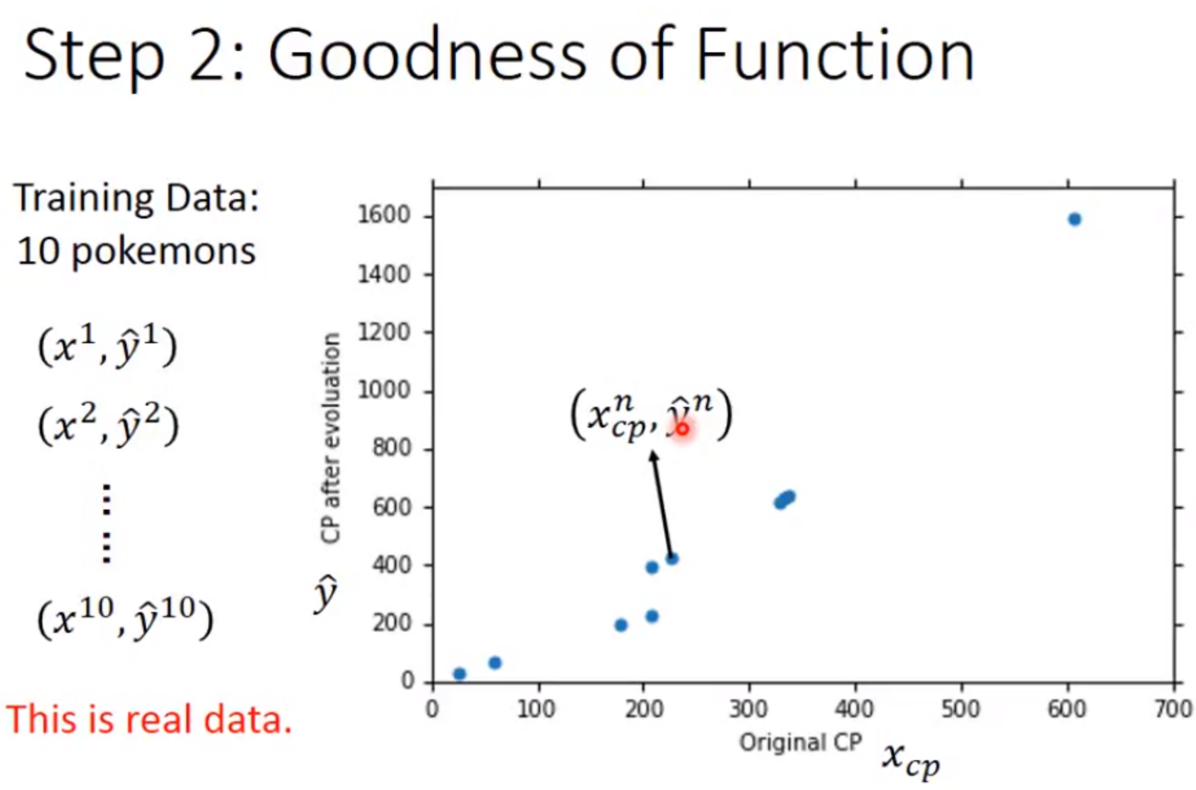
2. 样本确立

符号说明：

$$x^i \text{表示第} i \text{个训练样本的自变量}$$

$$y^i \text{表示第} i \text{个训练样本的标签}$$

$$x_j \text{表示自变量的第} j \text{个元素}$$



3. 构建损失函数

而在得到了训练数据集，以及确定了某些函数模型（function），就需要用损失函数（loss function）来衡量这些模型。　　我们假设所创建的模型为

$$y = b + w * x_{cp}$$

则损失函数可以设定为

$$L(f) = \sum i = 1^{10} \left(y^n - f(xcp^n)\right)^2$$

即通过标签y^n与函数$f(x_{cp}^n)$所预测出的值进行减法处理获取差异值。

得到所有样本的差异值的平方相加则为损失函数的值。

而L(f) 中的f表示的是函数模型集中的

$$y = b + w * x_{cp}$$

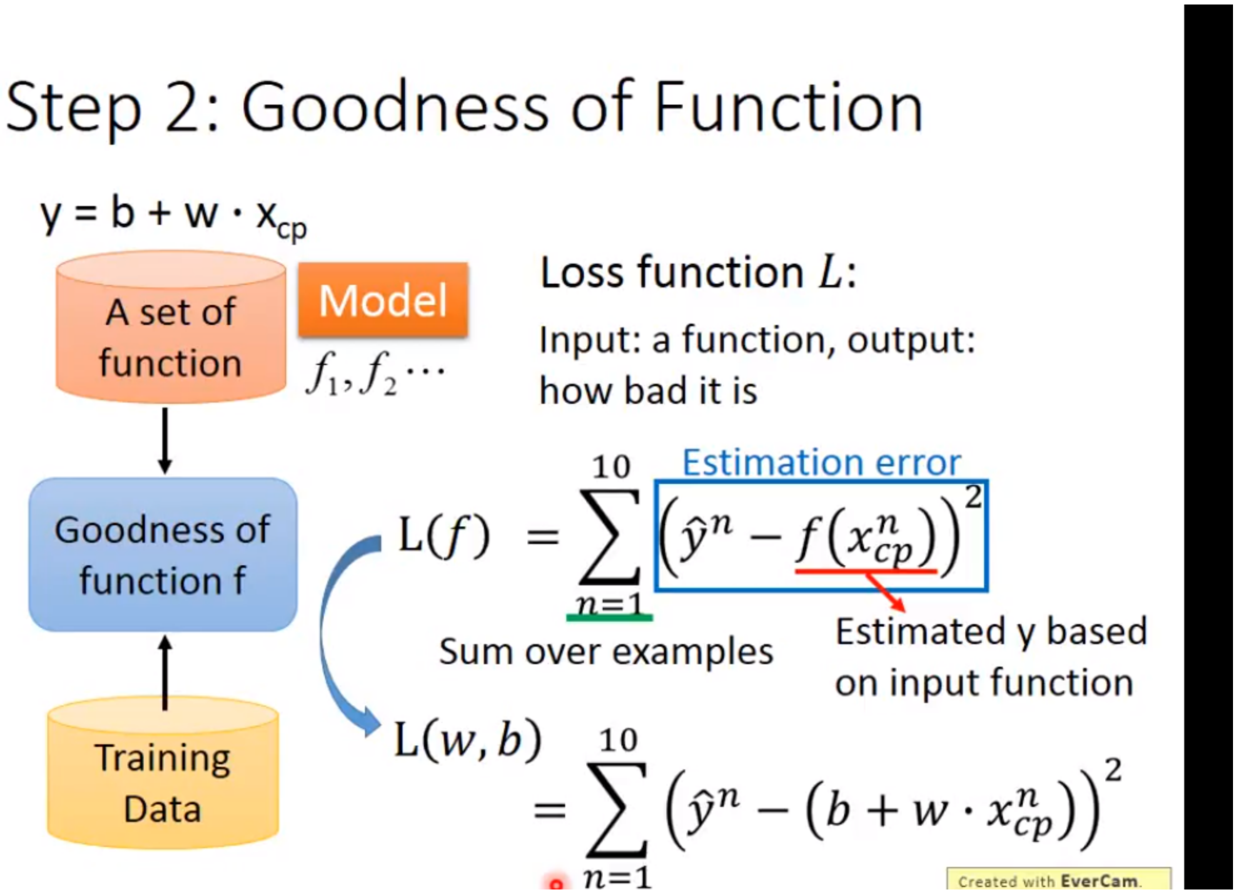
其中的变量有关的是w、b，因此可以将损失函数改动为$L(w, b)$。

则损失函数的公式为：

$$L(f) = \sum i = 1^{10} \left(y^n - (b + w * xcp^n)\right)^2$$

其中的变量有关的是w、b，因此可以将损失函数改动为L(w,b)。则损失函数的公式为：

$$L(f) = \sum i = 1^{10} \left(y^n - (b + w * xcp^n)\right)^2$$



4. 挑选最优函数模型

在创建了损失函数之后，我们按照损失函数对所有的模型进行测试 并确定不同参数下损失函数的值。根据这个值来挑选最优的函数模型。

已知损失函数确立后，最优的函数模型的损失函数值最小。 因此我们可以创建一个函数 f^* 来求得使损失函数值最小化的函数模型：

$$f^* = argmin_f L(f)$$

$$其中f表示的y = b + w * x_{cp}。$$

因此我们可以等价替换为

$$w^*, b^* = argmin_{(w,b)} L(w,b)$$

即求能将损失函数最小化的$w，b$的值。

$$w^*, b^* = argmin_{(w,b)} L(w,b)$$

$$= argmin(w,b) \sum i = 1^{10} \left(y^n - (b + w * x_{cp}^n)\right)^2$$

5. 梯度下降（Gradient Descent)

梯度下降的方法步骤：
（假设只有一个参数w）

- 随机挑选一个初值$$
- 计算

$$\frac{dL}{dw}|w=w^0$$
$$(对w^0设置一个初值)$$

  ,并且更迭

$$w^1 \leftarrow w^0 - \eta\frac{dL}{dw}|_{w=w^0}$$
$$（其中的\eta被称作学习率）$$

- 计算

$$\frac{dL}{dw}|w=w^1,$$
$$并且更迭w^2 \leftarrow w^1 - \eta\frac{dL}{dw}|w=w^1$$

- ......多次迭代
- 得到最终优化参数值

（假设有两个参数w,b）

- 随机挑选一个初值w^0,b^0
- 计算

$$\frac{\partial L}{\partial w}|w=w^0,b=b^0$$
$$\frac{\partial L}{\partial b}|w=w^0,b=b^0$$

  ,并且更迭

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

(其中的 η被称作学习率)

- 计算

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

  ,并且更迭

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

- ......多次迭代
  *得到最终优化参数值

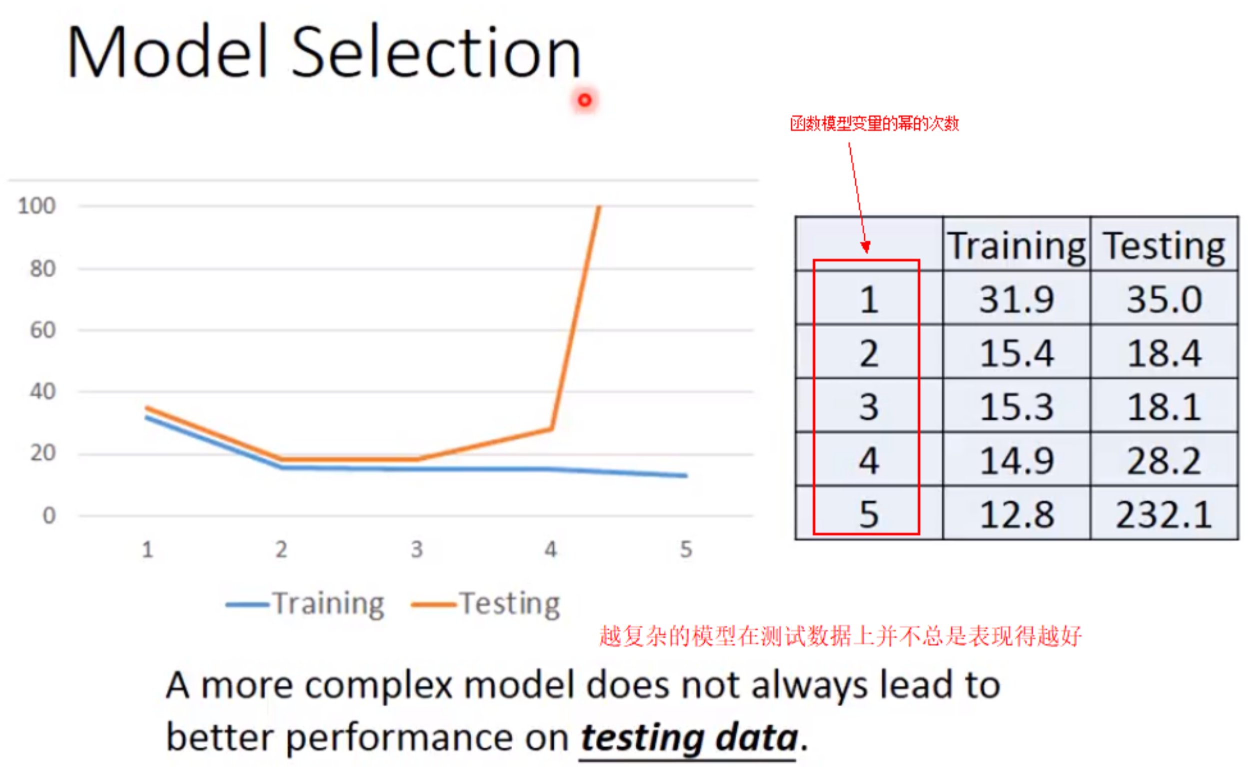
根据原式:

$$L(w,b) = \sum i = 1^{10}\left(y^n - (b + w * xcp^n)\right)^2$$

我们可以对其进行偏微分:

$$\frac{\partial L}{\partial w} = \sum i = 1^{10} 2\left(y^n - (b + w * xcp^n)\right) * (-x_{cp}^n)$$
$$\frac{\partial L}{\partial b} = \sum i = 1^{10} 2\left(y^n - (b + w * xcp^n)\right) * (-1)$$

而在不断地进行各种模型进行样本测试之后，可以发现如下的



因此我们发现当变量幂的次数为3时，对测试数据有更好的拟合效果。

5. 正则化

正则化表示如下：

$$L = \sum_{i=1} \left(y^n - (b + \sum w_i * x_i)\right)^2 + \sum (w_i)^2$$

通过 $\sum (w_i)^2$ 将 L 函数变得更平滑。

选用正则化的原因是：因为在大多数情况下更平滑的函数更趋于正确