拟合与优化

寻找正解

预备知识

课前

- Coursera吴恩达《机器学习》WEEK 2、3

https://www.coursera.org/learn/machine-learning

- 《数据挖掘导论》附录D、E

课后

- 网易公开课《机器学习》第2、3课

http://open.163.com/special/opencourse/machinelearning.html

拟合的本质

不确定到确定的自然规律

- 人类知识的演化

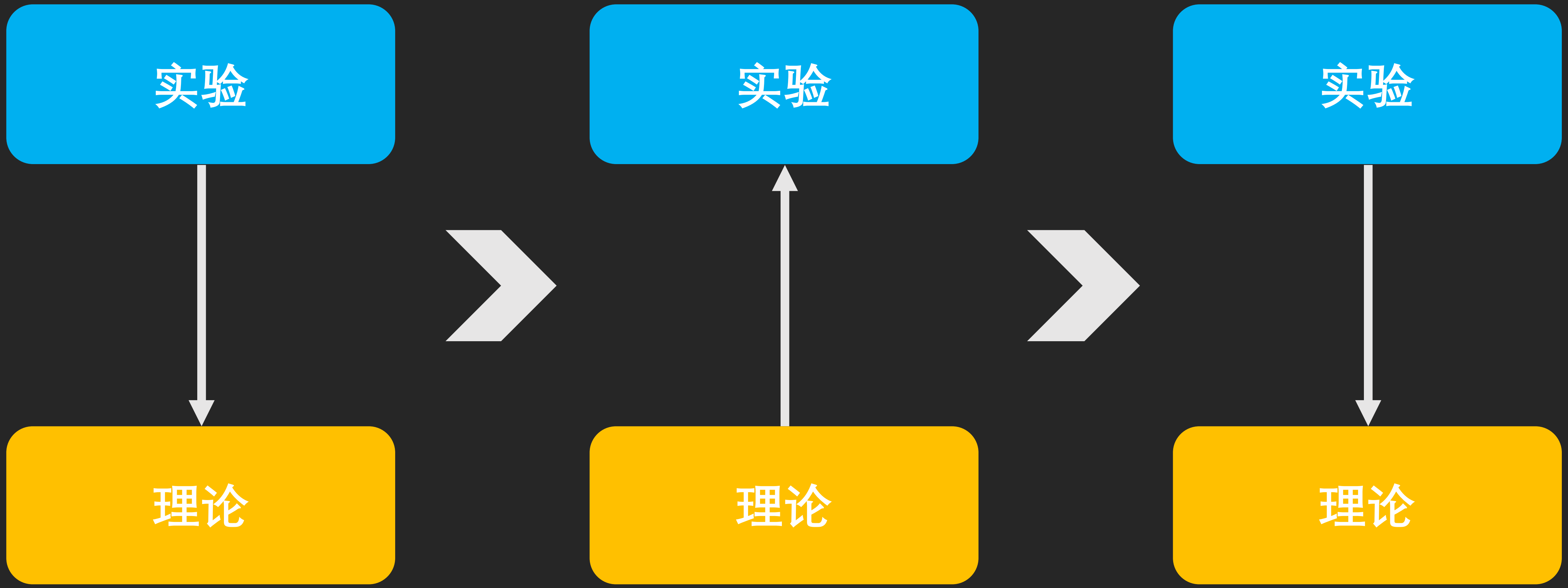
模糊 – 精确；经验性 – 理论基础

- 数据的噪声掩盖了正解——不确定性

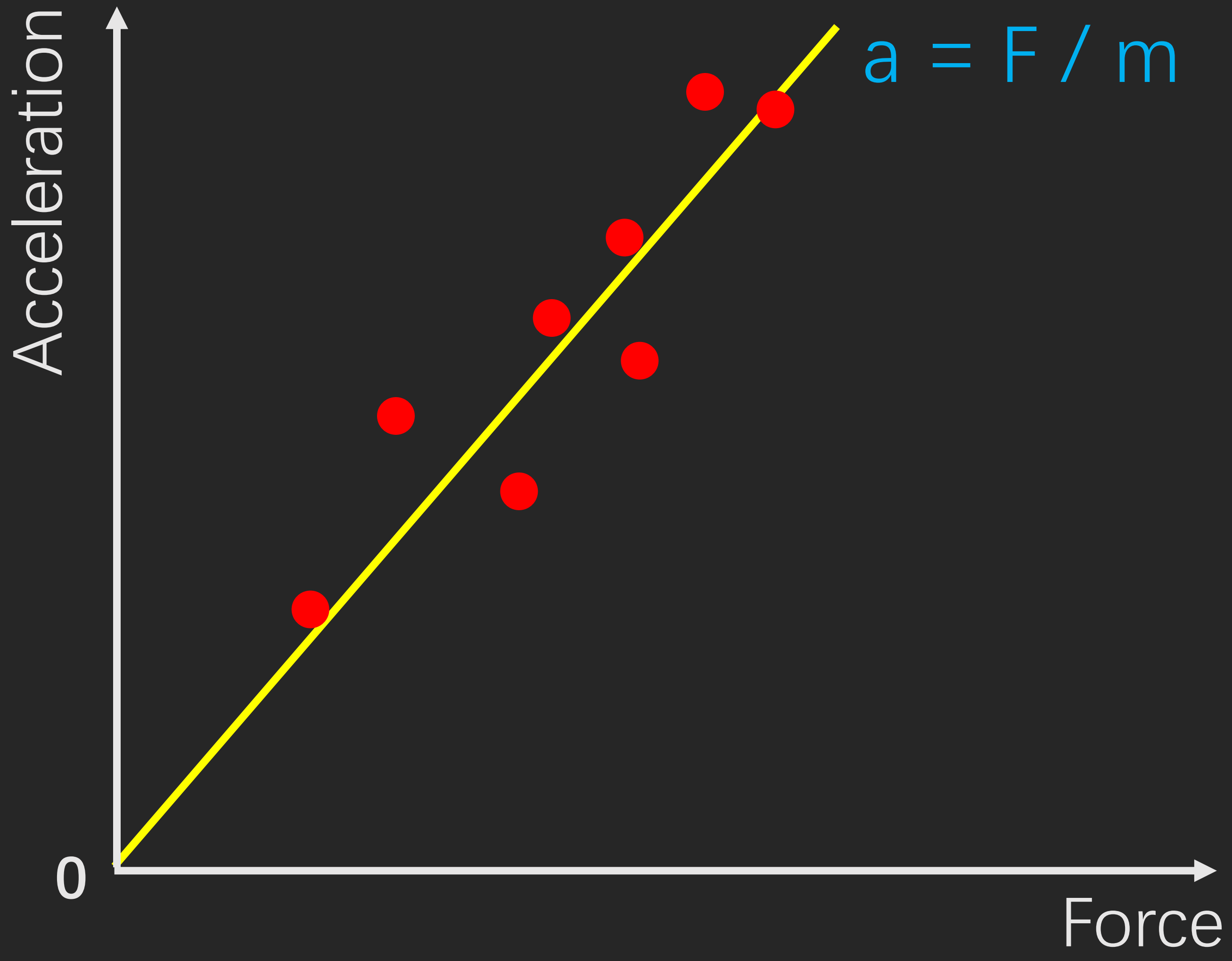
大数据让我们从噪声中挖掘出统计规律

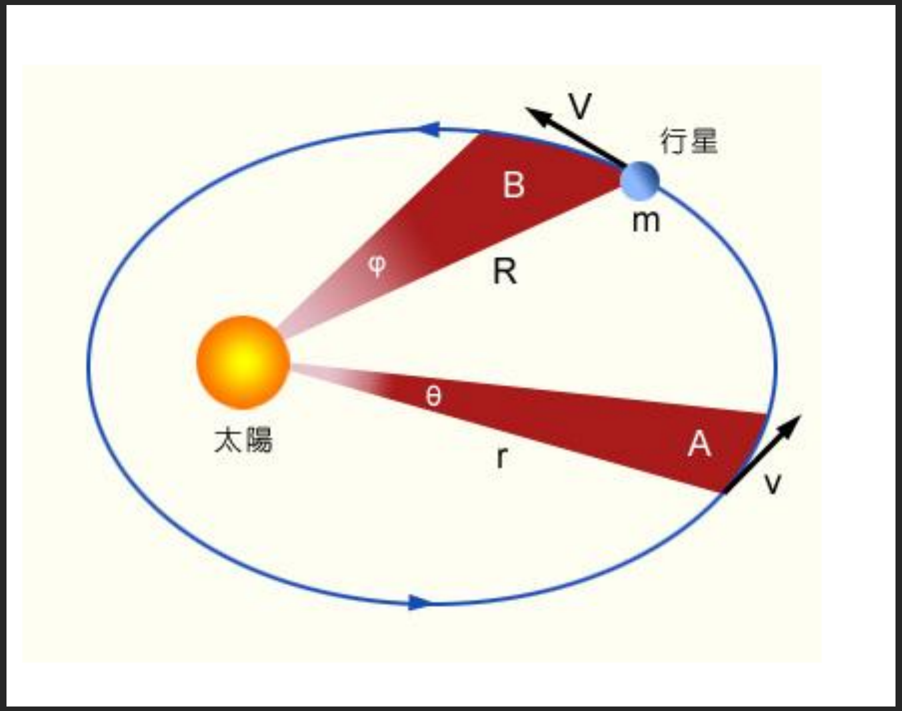
从而降低了不确定性

科学范式的转变



牛顿机与开普勒机





数学模型

线性成本函数：

$$J(\theta) = \sum \tfrac{1}{2} (y_i - \theta \cdot x_i)^2$$

非线性成本函数：

$$J(\theta) = \sum \tfrac{1}{2} (y_i - h_{\theta}(x_i))^2$$

目标：

$$\min \{Y\}$$

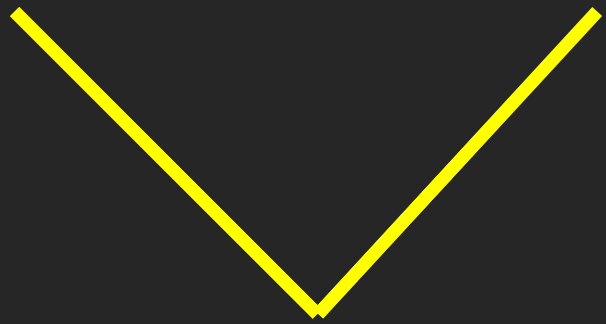
为什么成本函数不在一次元？

$J(\theta) = \sum \frac{1}{2} (y_i - \theta \cdot x_i)$

$J(\theta) = \sum \frac{1}{2} |y_i - \theta \cdot x_i|$

没有下限！

不可导！



线性拟合的闭合解

$$X = \begin{bmatrix} — (x^{(1)})^T — \\ — (x^{(2)})^T — \\ \vdots \\ — (x^{(m)})^T — \end{bmatrix} \quad \vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

$$J(\theta) \quad = \quad \frac{1}{2}(X\theta - \vec{y})^T(X\theta - \vec{y})$$

$$X\theta - \vec{y} \quad = \begin{bmatrix} (x^{(1)})^T\theta \\ \vdots \\ (x^{(m)})^T\theta \end{bmatrix} - \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

$$\boxed{\theta = (X^TX)^{-1}X^T\vec{y}.}$$

$$
\begin{aligned}
\nabla_\theta J(\theta) &= \nabla_\theta \frac{1}{2}(X\theta - \vec{y})^T(X\theta - \vec{y}) \\
&= \frac{1}{2}\nabla_\theta \left(\theta^T X^T X\theta - \theta^T X^T \vec{y} - \vec{y}^T X\theta + \vec{y}^T \vec{y}\right) \\
&= \frac{1}{2}\nabla_\theta \,\mathrm{tr}\left(\theta^T X^T X\theta - \theta^T X^T \vec{y} - \vec{y}^T X\theta + \vec{y}^T \vec{y}\right) \\
&= \frac{1}{2}\nabla_\theta \left(\mathrm{tr}\,\theta^T X^T X\theta - 2\mathrm{tr}\,\vec{y}^T X\theta\right) \\
&= \frac{1}{2}\left(X^T X\theta + X^T X\theta - 2X^T\vec{y}\right) \\
&= X^T X\theta - X^T\vec{y}
\end{aligned}
$$

优化方法

- 梯度下降

一阶梯度下降

- 牛顿法

二阶梯度下降

- 加入动量的梯度下降

加速收敛

梯度下降

$$\theta = \theta - \alpha \cdot \nabla_\theta J(\theta)$$

α – 学习速率

∇_θ – 梯度

对线性拟合问题有

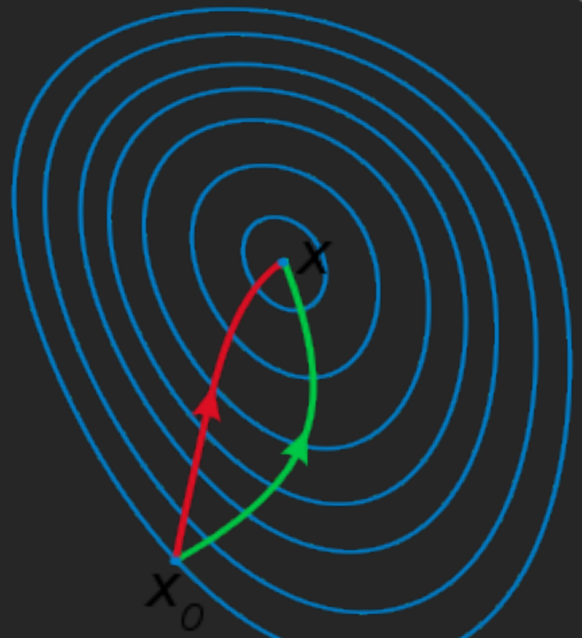
$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

牛顿法

$$\theta = \theta - \alpha \cdot H^{-1} \nabla_\theta J(\theta)$$

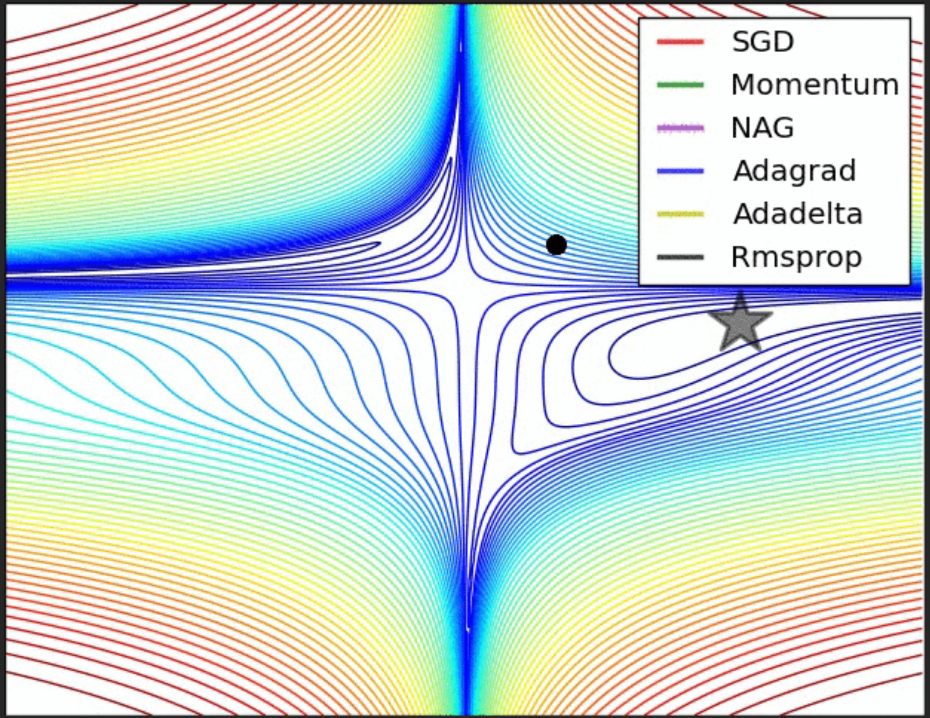
H – Hessian矩阵

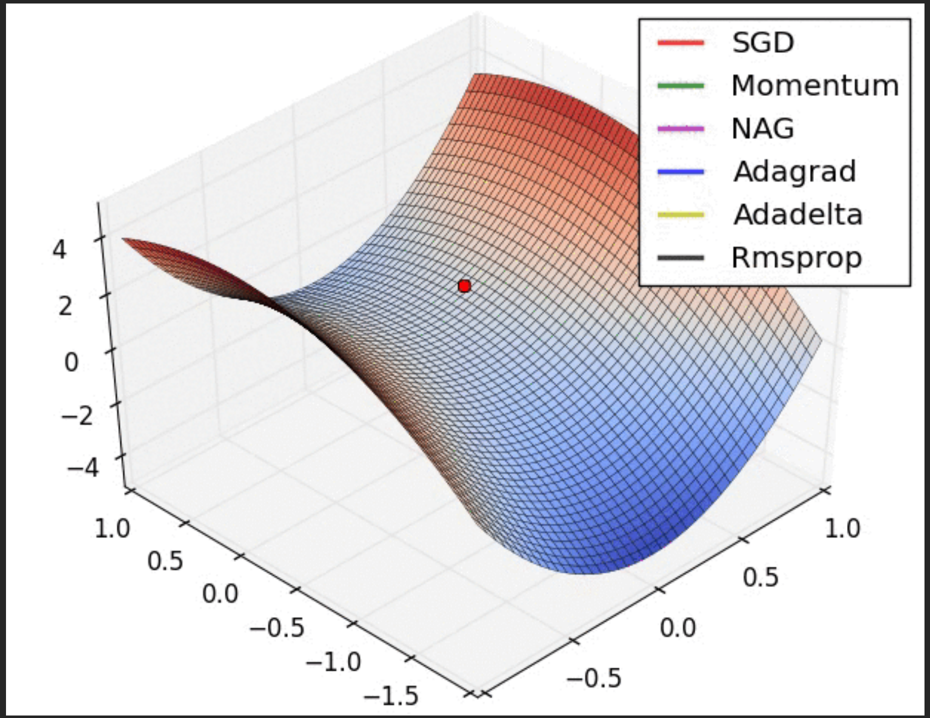
$$H_{ij} = \frac{\partial^2 \ell(\theta)}{\partial \theta_i \partial \theta_j}$$



动量的作用

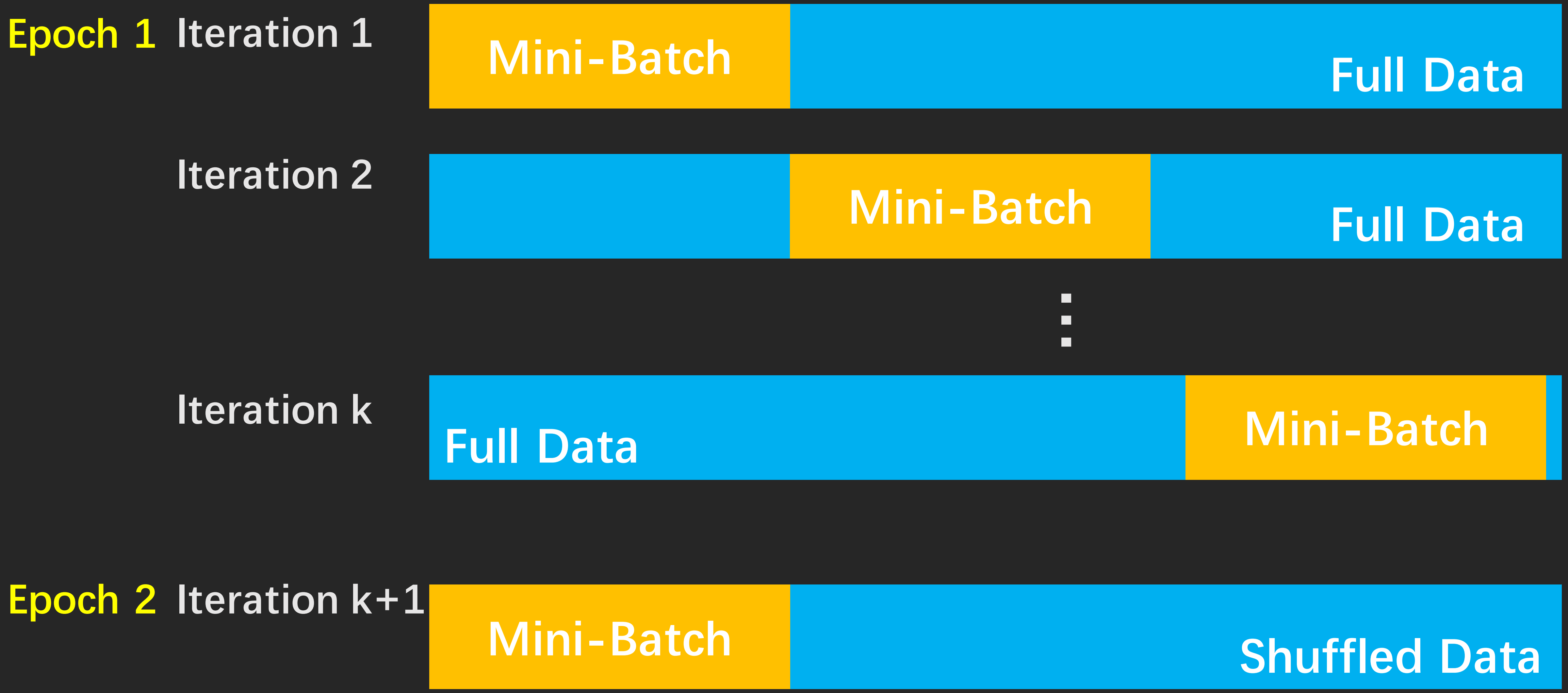
$$\Delta\theta_t = \gamma \cdot \Delta\theta_{t-1} + \alpha \cdot \nabla_\theta J(\theta)$$



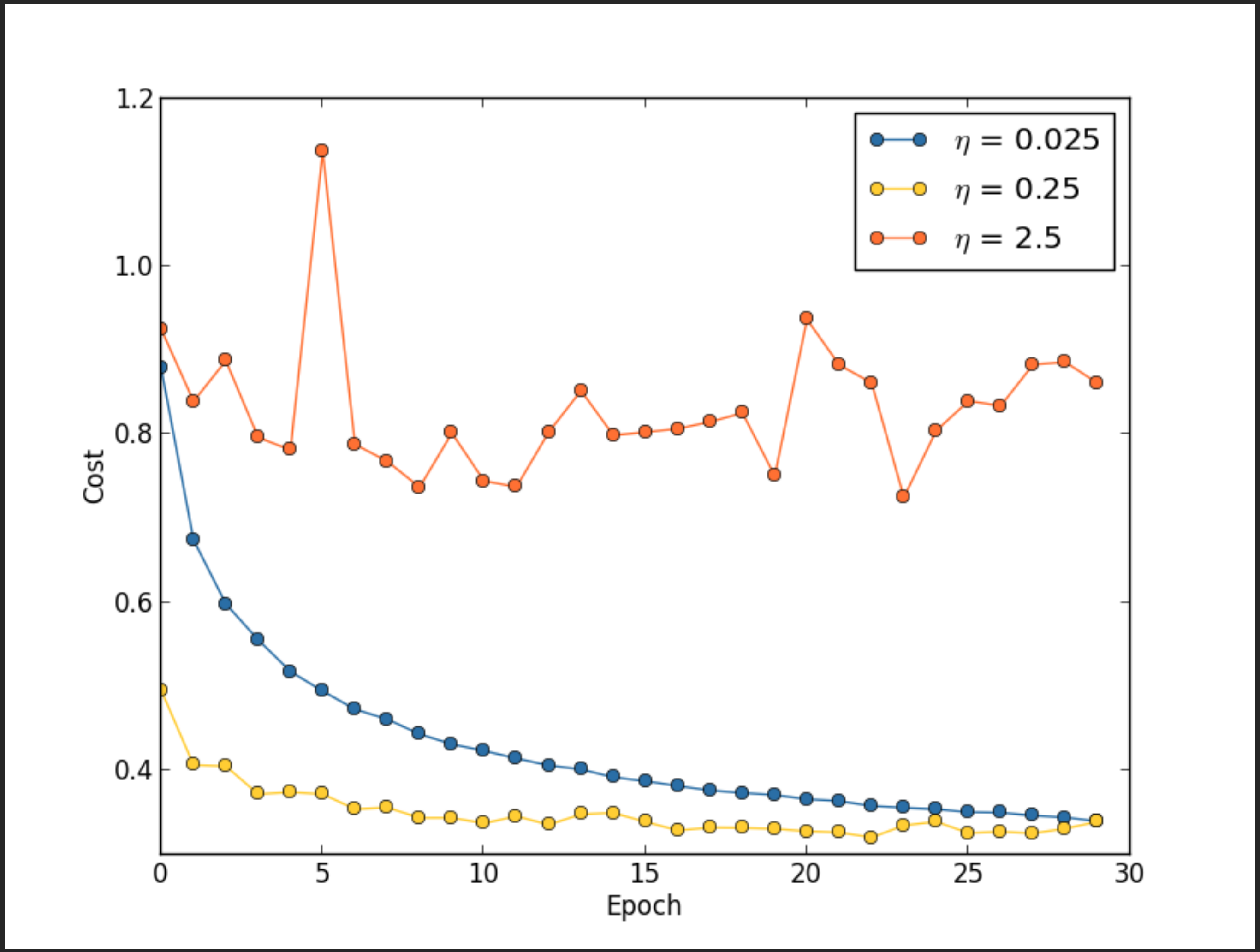


用多少数据优化参数？

Mini-Batch in Action



学习速率的重要性



学习速率的选择

- 人工调节学习速率

人工观察成本函数的变化（Human Intelligence）

- 可变的学习速率

例如：Every k iterations, set $\alpha = \alpha * 0.9$.

- Early Stopping

在正确的时间做正确的事情

Next Class – 拟合与优化（真枪实弹）

课前

- 学习Python的基本语法

https://docs.python.org/2/tutorial/index.html

- 了解Python的Scikit-Learn和TensorFlow机器学习框架

http://scikit-learn.org/

https://www.tensorflow.org/