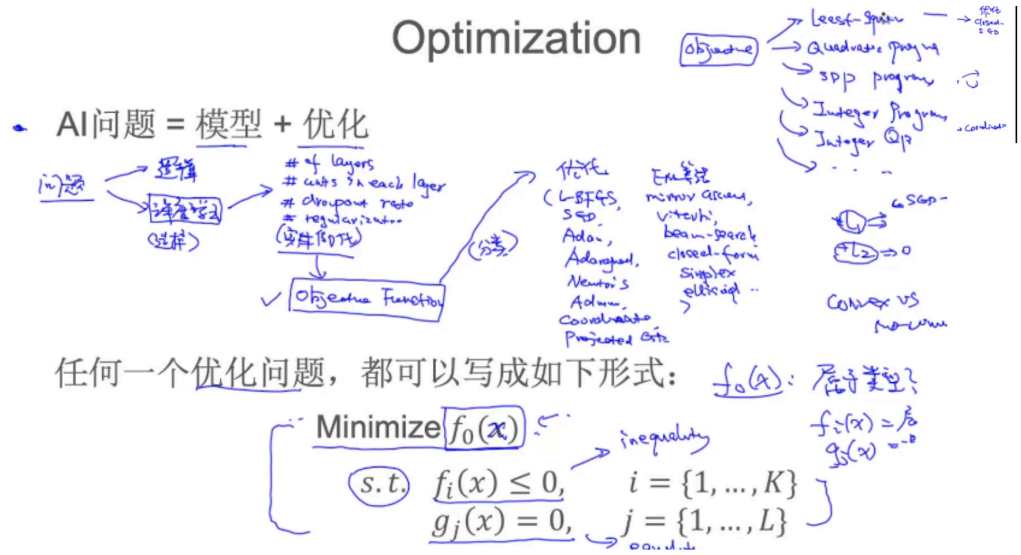
1. Optimization

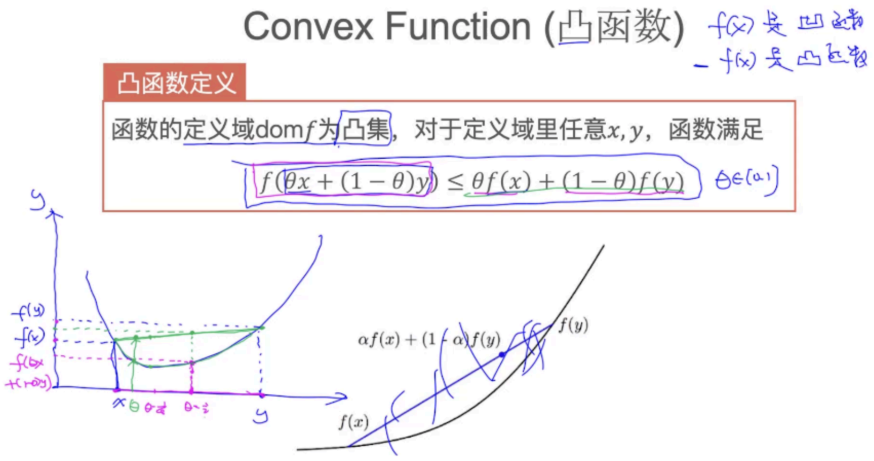


Optimization is the Core of Machine Learning
线性回归(Linear Regression)
逻辑回归(Logistic Regression
SVM(Support Vector Machine)
协同过滤(collaborative filtering)
K 均值(K-means)

1.1 categories

1)smooth / non-smooth
2)convex / non-convex
　　Convex 一定是全局最优解
　　Non-convex 是局部最优解，e.g.深度学习
3)discrete / continuous
　　Discrete 一般需要用到离散数学
4)constrained / non-constrained

1.2 convex function

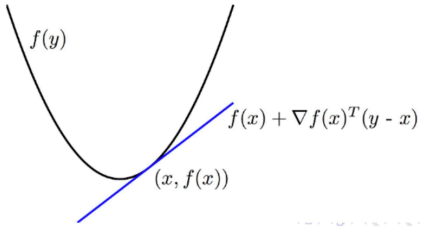


- 线性函数为凸/凹函数
- $\exp x, -\log x, x\log x$ 是凸函数
- 范数为凸函数
- $\frac{x^T x}{t}$为凸函数（x>0）

First Order Convexity Condition

假设f：$R^n \to R$是可导的（differentiable），则f 为凸函数，的当且仅当：　　$f(y) \geq f(x) + \nabla(x)^T(y-x)$

对于任意$x, y \in domf$



Second Order Convexity Condition

假设f：$R^n \to R$是两次可导的（twice differentiable），则f 为凸函数，当且仅当：

$\nabla^2 f(x) \geq 0$

对于任意$x, y \in domf$

scalar ≥ 0

matrix → PSD

(positive = semi-definite matrix) 正定矩阵

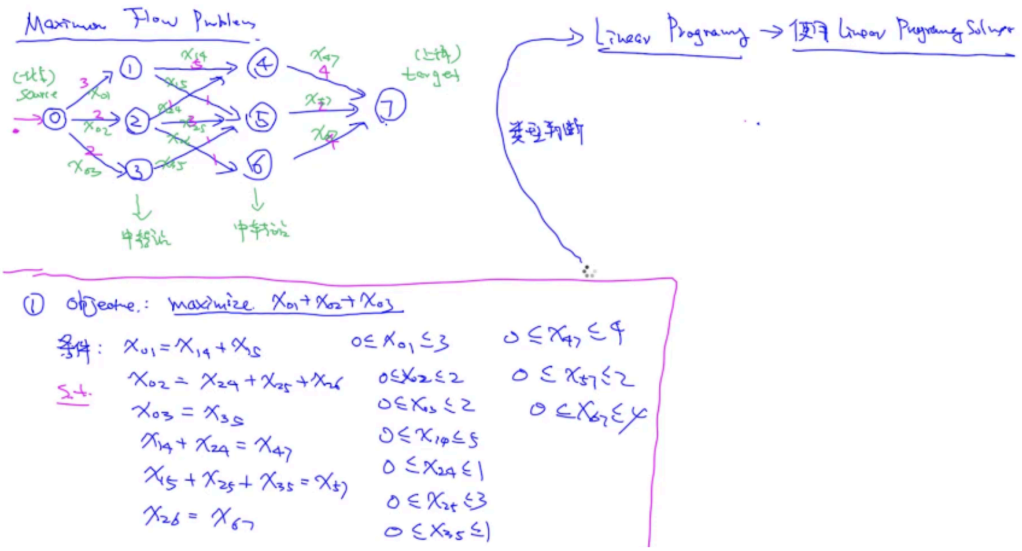
- 线性函数：$f(x) = b^T x + c$

$x_1, x_2 \Rightarrow f(x_1) = b^T x_1 + c \quad f(x_2) = b^T x_2 + c$

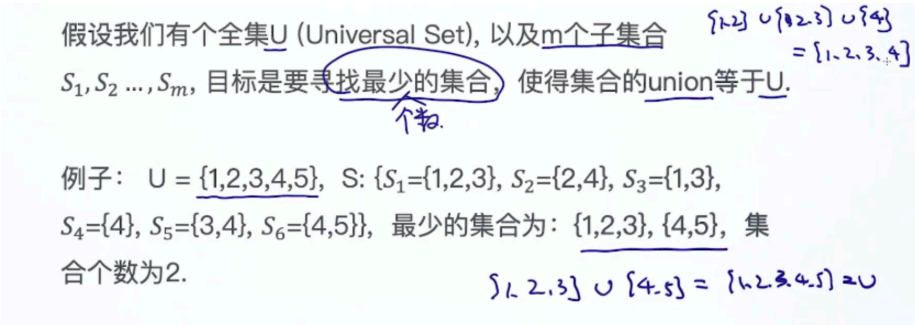
$$\Rightarrow b^T(\theta x_1 + (1-\theta)x_2) + c \leq \theta(b^T x_1 + c) + (1-\theta)(b^T x_2 + c)$$

$$\theta b^T x_1 + (1-\theta) b^T x_2 + c \leq \theta b^T x_1 + \theta c + c(1-\theta) b^T x_2 + (1-\theta)c$$

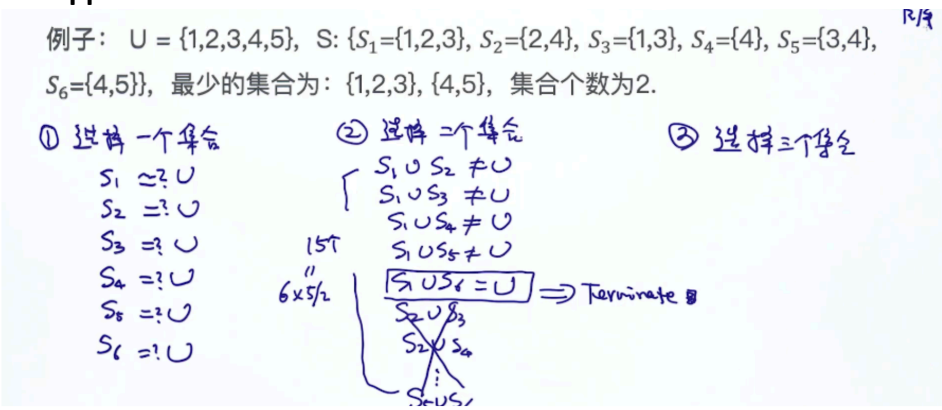
1.3 example-maximum flow problem



2. Set Cover Problem

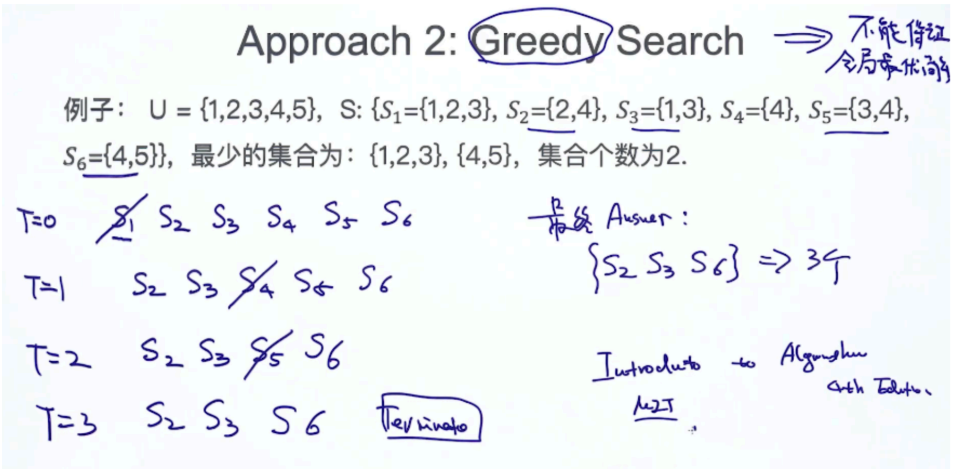


2.1 Approach 1: Exhaustive Search



全局最优解

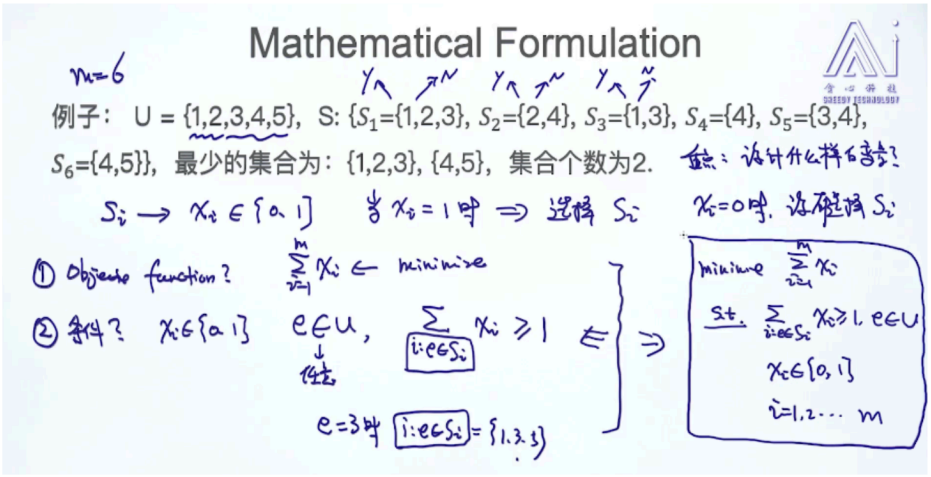
2.2 Approach 2: Greedy Search

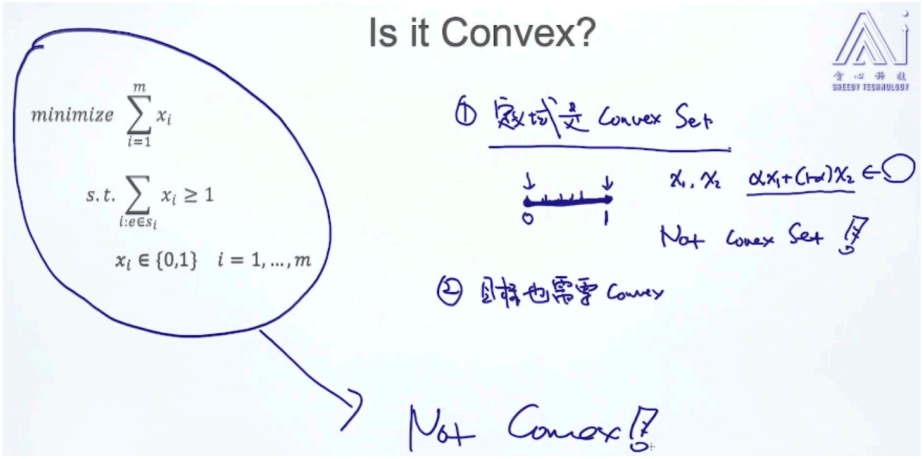


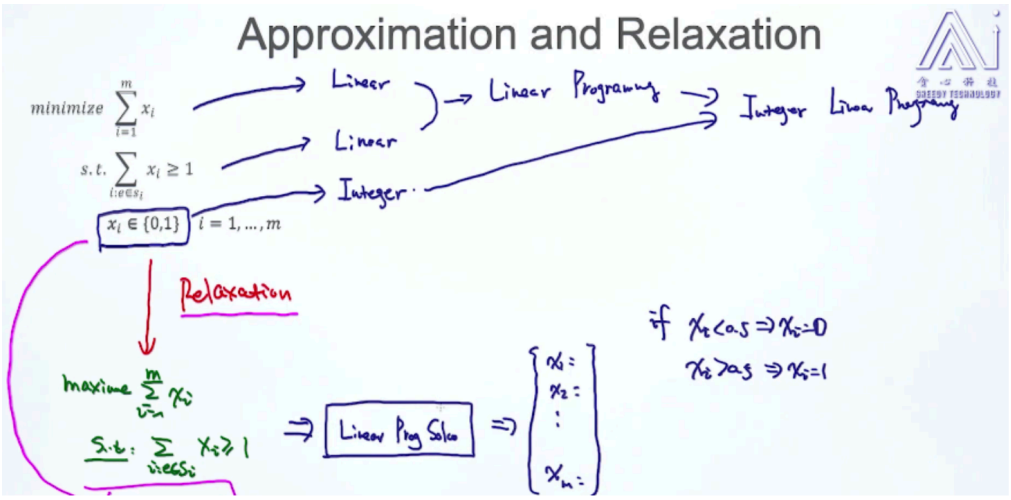
不能保证全局最优解

与穷举相比：时间高效，但是只是局部最优

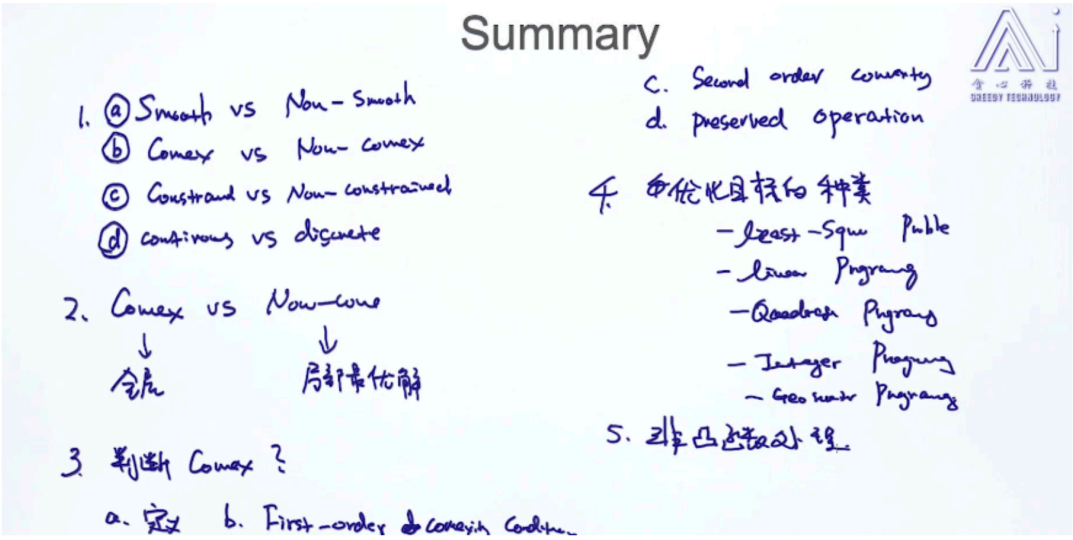
2.3 Approach3: Optimization







3. summary



4. 梯度下降法

4.1 Recall: Gradient Descent for Logistic Regression

逻辑回归是一个凸函数

$$p(y=1|x,w)=\frac{1}{1+e^{-w^Tx+b}}$$

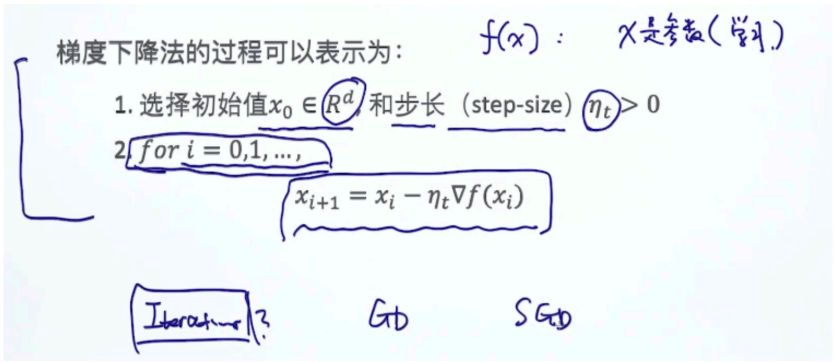
$$argmin_{w,b}-\sum_{i=1}^{n} y\log p(y=1|x,w)+(1-y)\log(1-p(y=1|x,w))]$$

$$w^{t+1}=w^t-\eta_t\sum_{i=1}^{n}[\sigma(w^Tx_i+b)-y_i]x_i$$

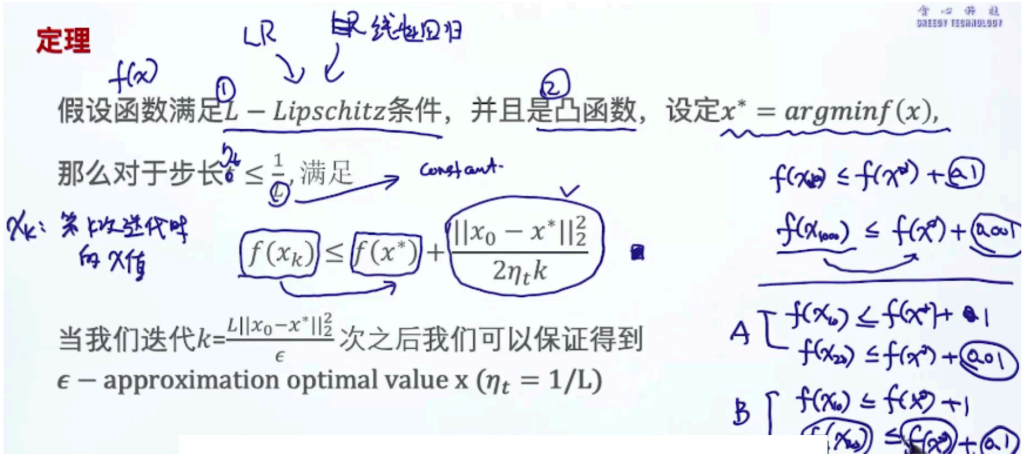
$$b^{t+1}=b^t-\eta_t\sum_{i=1}^{n}[\sigma(w^Tx_i+b)-y_i]$$

Gradient Descent

4.2 梯度下降表达

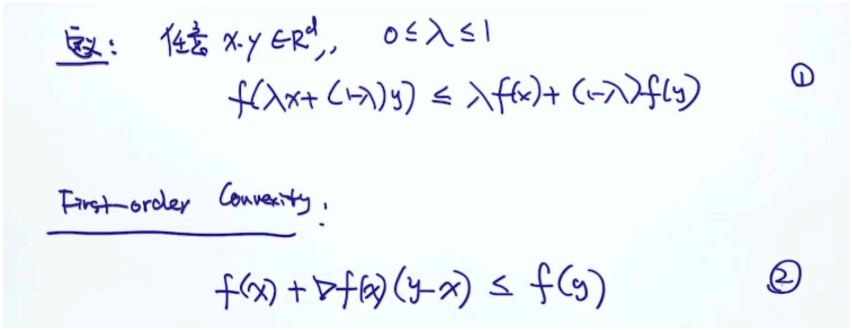


4.3 收敛分析 convergence analysis



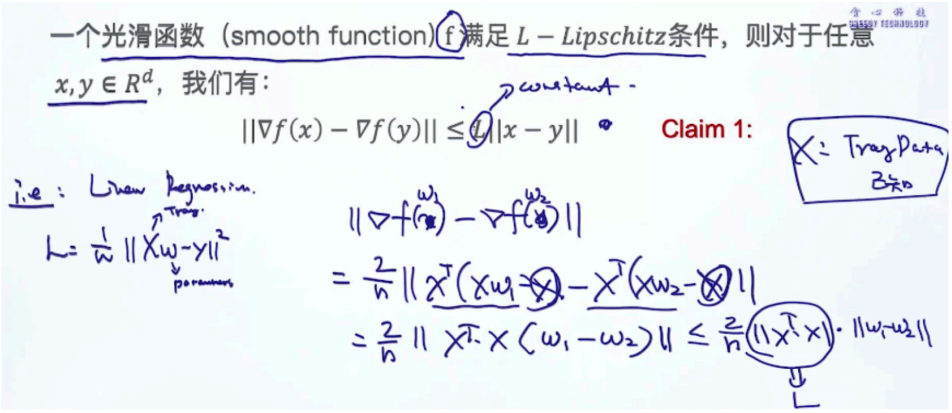
当 k 变大（迭代次数），不等式右边第二项会变小。那么 fxk 会慢慢趋向于 fx*。

4.4 凸函数性质回顾

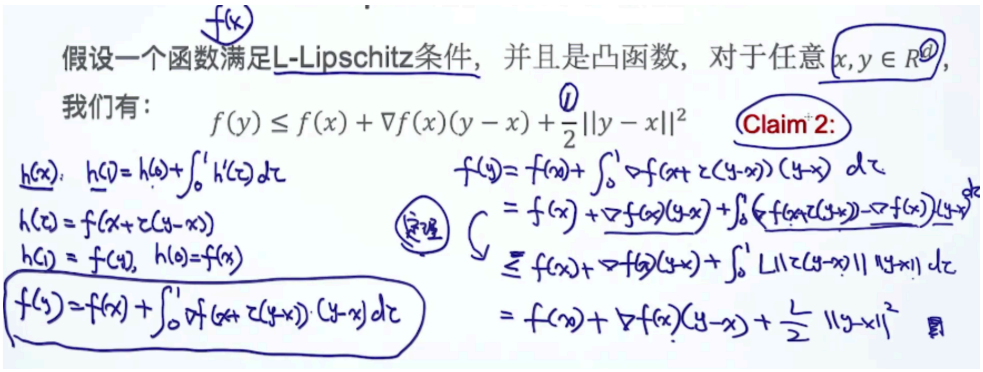


4.5 L-Lipschitz 条件及定理

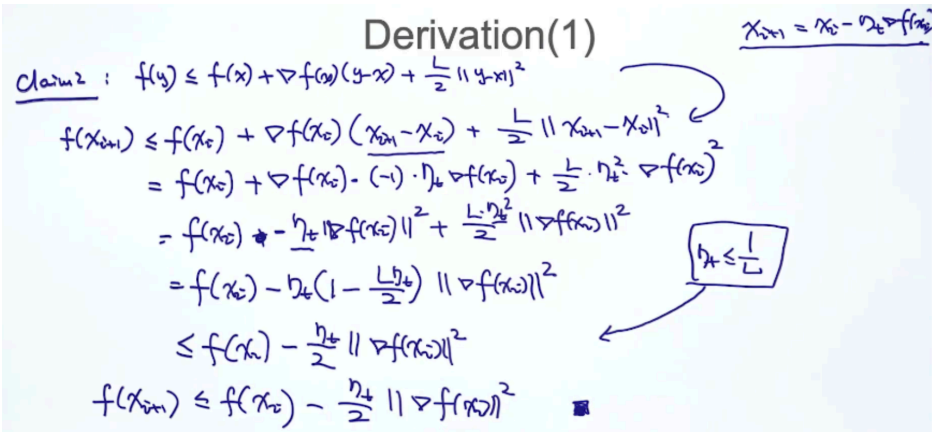
定理一



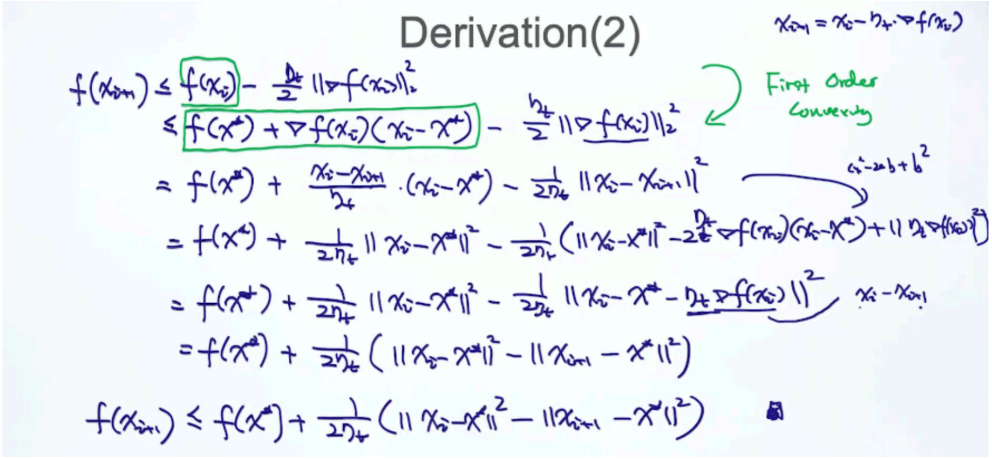
定理二



推导一



推导二



推导三