Gradient Descent

- Tip 1: Tuning your learning rates
  - $\eta^t = {}^\eta\!/_{\sqrt{t+1}}$
  - Giving different parameters different learning rates: AdaGrad
    - |First derivative|/(Second derivative)
- Tip 2: Stochastic Gradient Descent
- Tip 3: Feature Scaling
- Gradient Descent Theory: Multivariable Taylor Series
- More Limitation of GD: plateau, saddle point, local minima