Hyper-Parameters

Recap: Stochastic Gradient Descent

Optimizes model parameters θ

Failed to load file "/home/philkr/workspace/classes/deeplearning_v2/content/deep_networks/05_optimization/figures.ipynb.exec"

P

G

What About Other Parameters?

Hyper-Parameter

Any parameter not learned by SGD

Pseudocode: Stochastic Gradient Descent

```
θ ~ Init
for epoch in range(n):
  for (x, y) in dataset:
    J = ∇l(θ|x,y)
    θ = θ - ϵ * J.mT
```

Examples:

- Number of epochs n
- Learning rate ϵ
- Model architecture / size
- Loss function

How Do We Set Hyper-Parameters?

By hand (tuning)

- Using values that have worked for similar tasks
- Through iterative experimentation
- "Graduate Student Descent"