Lecture 11 Noisy Gradient Descent

Yu-Xiang Wang

COMPUTER SCIENCE
UC SANTA BARBARA
Computing. ReInvented.

Recap: Last lecture

- Convex empirical risk minimization
- Output perturbation
- Objective perturbation

Recap: Convex ERM and optimality conditions

- Data $(x_1, y_1), ..., (x_n, y_n) \in \mathcal{X} \times \mathcal{Y} = \mathcal{Z}$
- Convex ERM:

$$\min_{\theta \in \Theta \subset \mathbb{R}^d} \sum_{i=1}^{n} \ell(\theta, (x_i, y_i))$$

- Optimality condition: gradient = 0

- Assumptions: Lipschitzness, Smoothness

Recap: Output perturbation

- Stability of the output via regularization

- Privacy: from Gaussian mechanism
- Utility:
  - Last time: under smoothness (has a small error ☹)
  - Let’s do it again.

Recap: Utility of Output perturbation

- Smooth losses

- Lipschitz losses

Recap: Objective perturbation

- Algorithm $$\hat{\theta}^P = \operatorname*{argmin}_{\theta \in \Theta} L(\theta; D) + r(\theta) + \frac{\lambda}{2}||\theta||_2^2 + b^T\theta,$$

- Privacy analysis
    - For GLM
    - For General smooth learning problems

This lecture

- Utility analysis of objective perturbation

- Noisy Gradient Descent

- Privacy amplification by sampling and NoisySGD

Readings

- Chaudhuri et al. / Kifer et al. (continuing)

- Bassily et al. (2014) Private empirical risk minimization: Efficient algorithms and tight error bounds. In *FOCS*. https://arxiv.org/abs/1405.7085
  - For the NoisySGD algorithm
  - For NoisyGD just refer to this lecture note.

Utility analysis of objective perturbation

Checkpoint: Compare the excess empirical risk of Output/Objective Perturbation

	Lipschitz losses	Smooth losses	Smooth / Lipschitz GLM
Output Pert	$\frac{d^{1/4}L\|\theta^*\|\log(\frac{1}{\delta})^{1/4}}{n^{1/2}\epsilon^{1/2}}$	$\frac{d^{1/3}\beta^{1/3}L^{2/3}\|\theta^*\|^{4/3}\log(\frac{1}{\delta})^{1/3}}{n^{2/3}\epsilon^{2/3}}$	Same as left
ObjPert	Not applicable	$\frac{dL\|\theta^*\|\sqrt{\log(\frac{1}{\delta})}}{n\epsilon}$	$\frac{\sqrt{d}L\|\theta^*\|\sqrt{\log(\frac{1}{\delta})}}{n\epsilon}$

Lower order terms and dependence on β hidden.

- Normalized by 1/n to be consistent with prior tables.
- Non-private excess risk is on the order of $\sqrt{d/n}$
- Could be $O(d/n)$

What are not quite satisfactory?

- Require the loss to be twice differentiable
  - Convex losses need not be even differentiable

- We did not handle the constrained convex ERM

- They do not handle non-convex ERM problems, e.g., those that arise when optimizing deep neural networks

Gradient Descent

- Unconstrained, differentiable optimization problem

$$\min_x \ f(x)$$

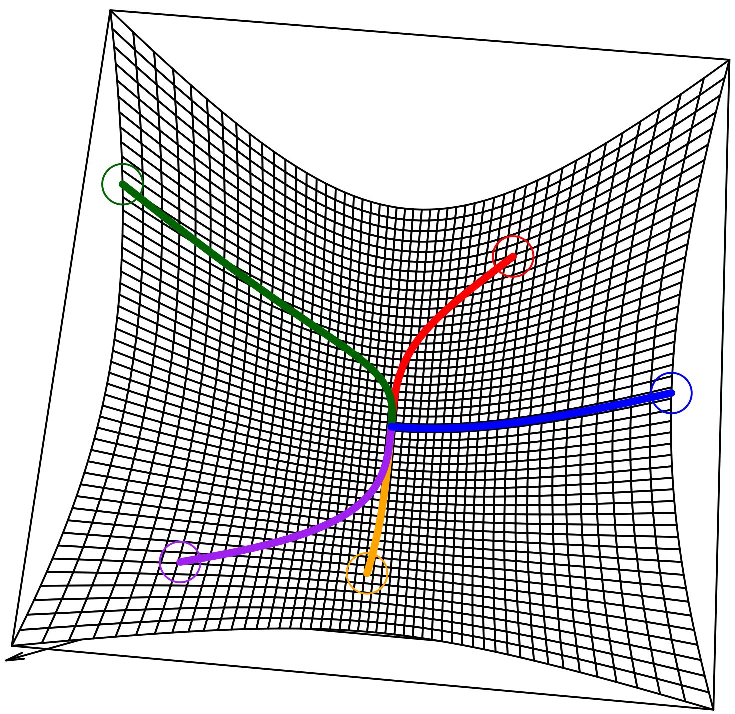
- The algorithm:

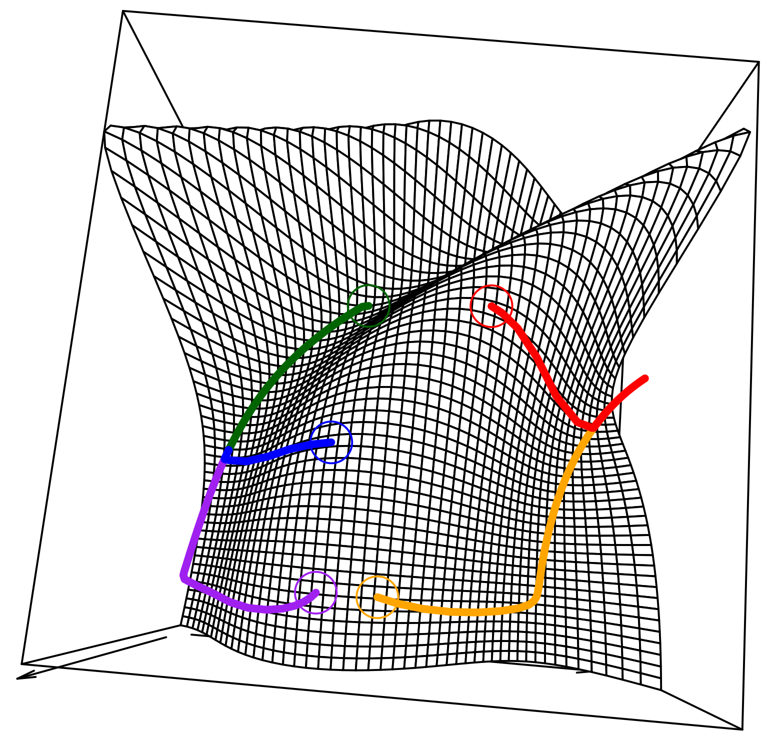
Gradient descent: choose initial point $x^{(0)} \in \mathbb{R}^n$, repeat:

$$x^{(k)} = x^{(k-1)} - t_k \cdot \nabla f(x^{(k-1)}), \quad k = 1, 2, 3, \ldots$$

Stop at some point

Gradient descent in convex problems vs nonconvex problems





Extensions of Gradient Descent

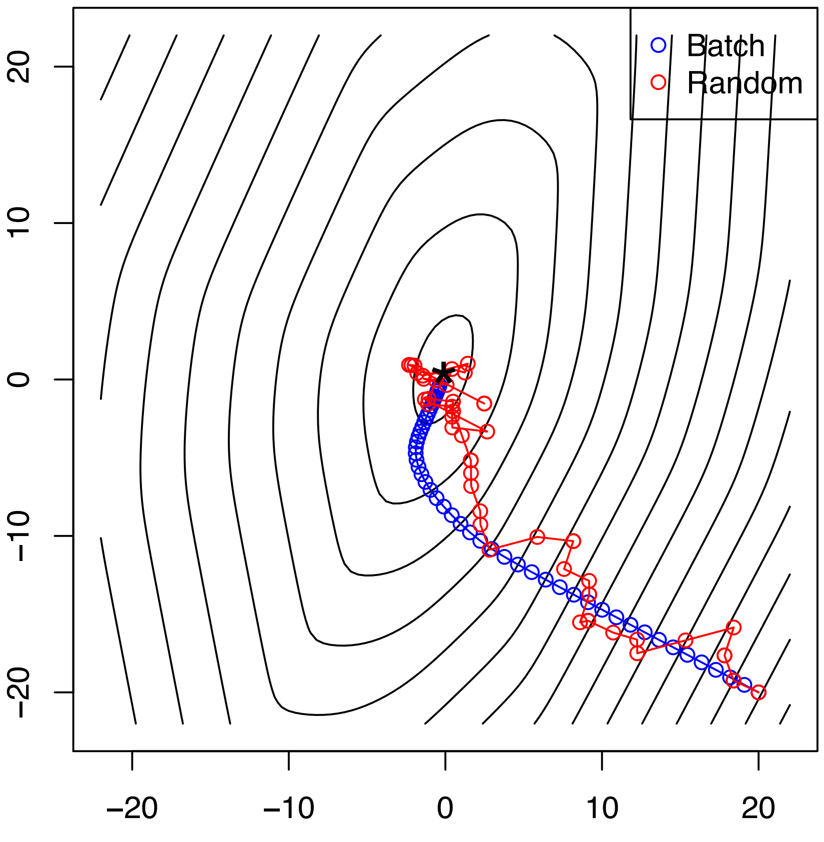
- Non-differentiable case: Subgradient descent
- Constrained case: Projected gradient descent
- Non-smooth penalty function: Proximal gradient descent
- Nonconvex cases: We give up theoretical guarantees but in practice it works (remarkably well)

Stochastic gradient descent

- Update rule:

$$\theta_{t+1} = \theta_t - \eta_t g_t$$

- Assumptions:



The convergence of GD and SGD

- GD in Smooth / convex problems
- GD in general convex problems
- SGD in general convex problems
- SGD in strongly convex problems
- Projected version

Convergence of stochastic gradient descent (in the smooth / nonconvex case)

- Descent Lemma

Convergence of stochastic gradient descent (in the smooth / nonconvex case)

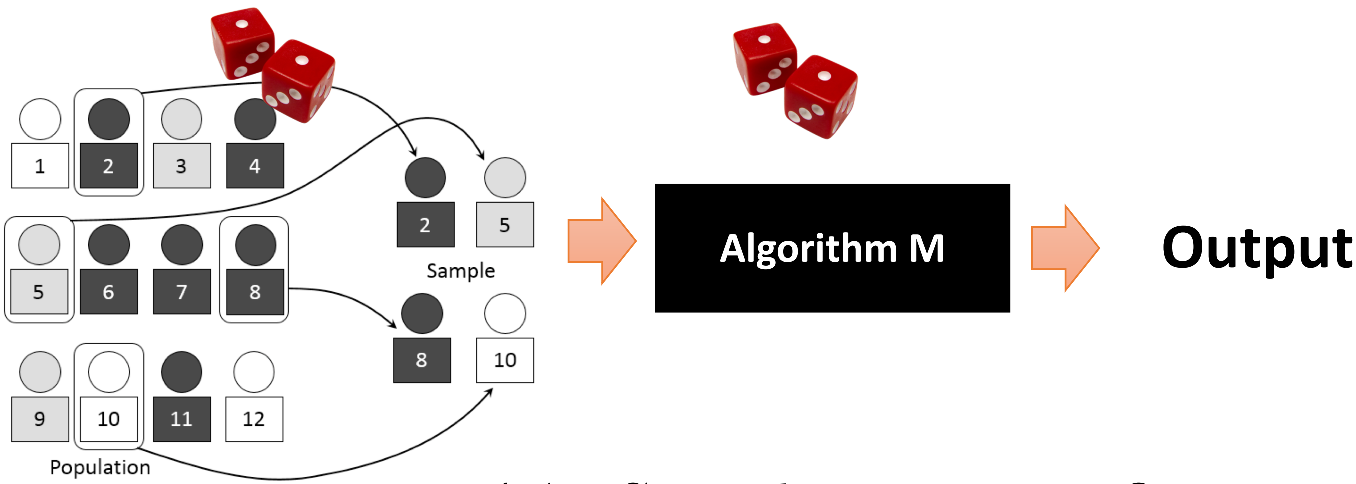
- Descent Lemma

Noisy Gradient Descent Mechanism for Convex ERM

- The algorithm:

- Privacy analysis:
  - A composition of T Gaussian mechanisms

Privacy Amplification by Sampling



$$\mathcal{M} \circ \text{Sample} : \text{Data} \to \text{Output}$$

Subsampling Lemma: If M obeys (Ɛ,δ)-DP, then M ∘ Subsample obeys that (Ɛ',δ')-DP with $\delta' = \gamma\delta$

$$\epsilon' = \log(1 + \gamma(e^{\epsilon} - 1)) = O(\gamma\epsilon)$$

Random subset sampling vs Poisson sampling

The Noisy Stochastic Gradient Descent Mechanism (NoisySGD)

- Privacy analysis:
  - A composition of T subsampled gaussian mechanism.

The Noisy Stochastic Gradient Descent Mechanism (NoisySGD)

- Utility analysis:
  - A composition of T subsampled gaussian mechanism.

Next lecture

- Differentially private deep learning

- Knowledge transfer model of private learning