Quiz 2

CSCI145/MATH166, Mike Izbicki

Definitions

$$z = (x, y) \sim P(z)$$

$$\hat{y} = f_w(x) = f(x; w)$$

$$\mathcal{F} = \text{a set of possible functions/models/hypotheses}$$

$$\ell(\hat{y}, y) = \text{loss function/cost of choosing wrong model}$$

$$E(f) = \mathbb{E}_z \ell(f(x), y)$$

$$E_n(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(x), y)$$

$$f^* = \arg\min_f E(f)$$

$$f^*_{\mathcal{F}} = \arg\min_{f \in \mathcal{F}} E(f)$$

$$f_n = \arg\min_{f \in \mathcal{F}} E_n(f)$$

$$\tilde{f}_n = f_{w_n} \text{where } w_n \text{ is given by one of the optimization formulas below}$$

$$\gamma_t = \text{step size of an optimization algorithm at iteration } t$$

Formulas

Formula for gradient descent (GD):

$$w_{t+1} =$$

Formula for stochastic gradient descent (SGD):

$$w_{t+1} =$$

The different types of errors

error type	definition	$n \uparrow$	$d \uparrow$	$\lvert\mathcal{F}\rvert \uparrow$	compute$\uparrow$
bayes error					
approximation error					
estimation error					
optimization error					
generalization error					

Rate of the optimization error

You may assume that the estimation error grows as $O(n^{-1/2})$.

Action	SGD	GD	2GD
iterations to accuracy ρ	$\frac{1}{\rho}$	$\log \frac{1}{\rho}$	$\log \log \frac{1}{\rho}$
time per iteration			
time to accuracy ρ			
time to est. error = opt. error = $\mathcal{E}$			

Which algorithm is typically fastest in small scale learning?

Which algorithm is typically fastest in large scale learning?