Title: Supplementary material for computer practical class 1

Author: Georgios Karagiannis (georgios.karagiannis@durham.ac.uk) January 21, 2024

The goal (roughly speaking...)

Consider a set of examples $\mathcal{S} = \{z_1, ..., z_n\}$ of size n, with $z_i \overset{\text{IID}}{\sim} g$.
Consider functions

$$R(w) = \mathrm{E}_{z\sim g}\left(\ell(w, z)\right), \qquad \text{and} \qquad R_{\mathcal{S}}(w) = \frac{1}{n}\sum_{i=1}^{n} \ell(w, z_i)$$

where $w \in \mathcal{W} \subset \mathbb{R}^d$ and $\ell(\cdot, \cdot)$ is a loss function.
Find minimizers in the following problems

$$w^* = \arg\min_w \{R(w)\} = \arg\min_w \{\mathrm{E}_{z\sim g}(\ell(w, z))\}$$

$$w^{**} = \arg\min_w \{R_{\mathcal{S}}(w)\} = \arg\min_w \left\{\frac{1}{n}\sum_{i=1}^{n} \ell(w, z_i)\right\}$$

Both of the following algorithms are often used to approximate w^* and w^{**}

Gradient descent algorithm

Provide:

- A seed (initial value) $w^{(0)}$
- Learning rate $\eta_t : \mathbb{N} \rightarrow \mathbb{R}_+$; a positive non-increasing function such as $\eta_t \rightarrow 0$. E.g.
  - $\eta_t = t_0/t$, with $t_0 > 0$ user specified
  - $\eta_t =$ something retatively tiny

For $t = 1, 2, 3, ...$ iterate:

1. compute
$$w^{(t+1)} = w^{(t)} - \eta_t \nabla R_{\mathcal{S}}\left(w^{(t)}\right)$$
2. terminate if a termination criterion is satisfied, e.g.
$$\text{If } \ t \geq T_{\max} \ \text{ then } \ \text{STOP}$$

Return $\bar{w} = \text{mean}\left\{w^{(t)}\right\}$

Stochastic gradient descent algorithm

Provide:

- A seed (initial value) $w^{(0)}$
- Learning rate $\eta_t : \mathbb{N} \to \mathbb{R}_+$; a positive non-increasing function such as $\eta_t \to 0$. E.g.
  - $\eta_t = t_0/t$, with $t_0 > 0$ user specified
  - $\eta_t =$ something retatively tiny
- An integer value for $m \in \{1, ..., n\}$, i.e. $m << n$.

For $t = 1, 2, 3, ...$ iterate:

1. collect a random sub-sample $\tilde{\mathcal{S}} = \left\{ \tilde{z}_j^{(t)}; j = 1, ..., m \right\}$ of size m with or without replacement from the complete data-set $\mathcal{S}$
2. compute
$$w^{(t+1)} = w^{(t)} - \eta_t \nabla R_{\tilde{\mathcal{S}}} \left(w^{(t)} \right)$$
where $R_{\tilde{\mathcal{S}}}(\cdot) = \frac{1}{m} \sum_{i=1}^{m} \ell(\cdot, \tilde{z}_i)$.
3. terminate if a termination criterion is satisfied, e.g.
$$\text{If } \; t \geq T_{\max} \; \text{ then } \; \text{STOP}$$

Return $\bar{w} = \text{mean} \left\{ w^{(t)} \right\}$