15-859 Algorithms for Big Data Assignment 1

Fan Pu Zeng
fzeng@andrew.cmu.edu

1: Scratcy Scratch

$$\begin{aligned}
-\nabla\mathcal{L}(\mathbf{w}) &= -\nabla \sum_{n=1}^{N} \ell\left(y_n \mathbf{w}^\top \mathbf{x}_n\right) \\
&= -\nabla \sum_{n=1}^{N} \exp\left(-y_n \mathbf{w}^\top \mathbf{x}_n\right) \\
&= \sum_{n=1}^{N} \exp\left(-\mathbf{w}(t)^\top \mathbf{x}_n\right) \mathbf{x}_n
\end{aligned}$$

$$\begin{aligned}
-\nabla\mathcal{L}(\mathbf{w}) &= \sum_{n=1}^{N} \exp\left(-\mathbf{w}(t)^\top \mathbf{x}_n\right) \mathbf{x}_n \\
&= \sum_{n=1}^{N} \exp\left(-g(t)\mathbf{w}_\infty^\top \mathbf{x}_n\right) \exp\left(-\boldsymbol{\rho}(t)^\top \mathbf{x}_n\right) \mathbf{x}_n
\end{aligned}$$

$$\hat{\mathbf{w}} = \sum_{n=1}^{N} \alpha_n \mathbf{x}_n \tag{1}$$

$$\text{such that} \qquad \forall n \left(\alpha_n \geq 0 \text{ and } \hat{\mathbf{w}}^\top \mathbf{x}_n = 1\right) \text{ OR } \left(\alpha_n = 0 \text{ and } \hat{\mathbf{w}}^\top \mathbf{x}_n > 1\right) \tag{2}$$

$$\begin{aligned}
q_m^\top k_n &= \left(\mathbf{W}_q \left(x_m + p_m\right)\right)^\top \left(\mathbf{W}_k \left(x_n + p_n\right)\right) \\
&= x_m^\top \mathbf{W}_q^\top \mathbf{W}_k x_n + x_m^\top \mathbf{W}_q^\top \mathbf{W}_k p_n \\
&+ p_m^\top \mathbf{W}_q^\top \mathbf{W}_k x_n + p_m^\top \mathbf{W}_q^\top \mathbf{W}_k p_n
\end{aligned}$$