JHU Engineering for Professionals
Applied and Computational Mathematics
Data Mining: 625.740

Module 7 Homework Solutions

1. $f = (\mathbf{x}-\mathbf{x_q})^{\mathrm{T}}(\mathbf{x}-\mathbf{x_q}) + \alpha(\mathbf{w}^{\mathrm{T}}\mathbf{x_q}+w_0)$,
   $0 = \frac{\partial f}{\partial \mathbf{x_q}} = 2(\mathbf{x_q}-\mathbf{x}) + \alpha\mathbf{w}, \quad (*)$
   $0 = \frac{\partial f}{\partial \alpha} = \mathbf{w}^{\mathrm{T}}\mathbf{x_q} + w_0, \quad (**)$
   Solve for $\mathbf{x_q}$ in $(*)$ and substitute into $(**)$:
   $-\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0 = 0$,
   $\frac{\alpha}{2} = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}+w_0}{\mathbf{w}^{\mathrm{T}}\mathbf{w}} = \frac{g(\mathbf{x})}{|\mathbf{w}|^2}$.

   (a) From $(*)$: $\mathbf{x}-\mathbf{x_q} = \frac{\alpha}{2}\mathbf{w}$, so $|\mathbf{x}-\mathbf{x_q}| = \frac{|g(\mathbf{x})|}{|\mathbf{w}|}$.

   (b) $\mathbf{x}_p = \mathbf{x_q} = \mathbf{x} - \frac{\alpha}{2}\mathbf{w} = \mathbf{x} - \frac{g(\mathbf{x})}{|\mathbf{w}|^2}\mathbf{w}$.

2. Because Q is positive definite, all its eigenvalues are strictly positive and thus Q^{-1} exists and is positive definite. Similarly, Q^T and Q^{-T} are positive definite. The sum of positive definite matrices is positive definite and therefore invertible. Thus $P = \left[Q^{-1}+Q^{-T}\right]^{-1}$ exists.
   $0 = \frac{\partial}{\partial \mathbf{x}}\sum_{k=1}^n(\mathbf{x_k}-\mathbf{x})^TQ^{-1}(\mathbf{x_k}-\mathbf{x}) = \sum_{k=1}^n[Q^{-1}+Q^{-T}](\mathbf{x}-\mathbf{x_k}) = [Q^{-1}+Q^{-T}]\sum_{k=1}^n(\mathbf{x}-\mathbf{x_k})$.
   Left multiply by P: $\sum_{k=1}^n \mathbf{x} = \sum_{k=1}^n \mathbf{x}_k$, $\Rightarrow \mathbf{x} = \overline{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^n \mathbf{x}_k$.
   $\frac{1}{n}\frac{\partial^2}{\partial \mathbf{x}\partial \mathbf{x}^T}\sum_{k=1}^n(\mathbf{x_k}-\mathbf{x})^TQ^{-1}(\mathbf{x_k}-\mathbf{x}) = Q^{-1}+Q^{-T} > 0$, so we are at a minimum.

3. Without loss of generality, we show $g_i(\mathbf{x}_1) > 0$ & $g_i(\mathbf{x}_2) > 0 \Rightarrow g_i(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) > 0$.
   $g_i(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) = \mathbf{w}_i^{\mathrm{T}}[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] + w_{i0} = \lambda[\mathbf{w}_i^{\mathrm{T}}x_1 + w_{i0}] + (1-\lambda)[\mathbf{w}_i^{\mathrm{T}}\mathbf{x}_2 + w_{i0}]$
   $= \lambda g_i(\mathbf{x}_1) + (1-\lambda)g_i(\mathbf{x}_2) > 0$.

4. $\frac{\partial}{\partial a_j}\sum(a_jy_j-b)^2 = 2y_j\sum(a_jy_j-b) \Rightarrow \nabla_a(\mathbf{a}^T\mathbf{y}-b)^2 = 2(\mathbf{a}^T\mathbf{y}-b)\mathbf{y}$.
   $\frac{\partial^2}{\partial a_j\partial a_i}\sum_j(a_jy_j-b)^2 = 2\frac{\partial}{\partial a_i}\left[y_j\sum_j(a_jy_j+b)\right] = \begin{cases} 2y_i^2, & i=j, \\ 2y_iy_j, & i\neq j. \end{cases}$
   $D = \frac{\partial^2 J}{\partial \mathbf{a}\partial \mathbf{a}^T} = 2\mathbf{y}_1\mathbf{y_1}^T$.

   Recall that the optimal ρ_k is

$$\rho_k = \frac{||\nabla J||^2}{\nabla J^T D \nabla J} = \frac{1}{2\mathbf{y}_1^T\mathbf{y}_1}.$$

$$\mathbf{a}_{k+1} = \mathbf{a}_k - \rho_k\nabla J(\mathbf{a}_k) = \mathbf{a}_k - \frac{\mathbf{a}_k^T\mathbf{y}_1 - b}{\mathbf{y}_1^T\mathbf{y}_1}\mathbf{y}_1.$$

5. Rearranging,

$$y_i \sum_j \exp(a_j) = \exp(a_i).$$

Implicitly differentiating with respect to a_j yields

$$\frac{\partial y_i}{\partial a_j}\left(\sum_j \exp(a_j)\right) + \exp(a_j) y_i = \delta_{ij}\exp(a_j).$$

$$\frac{\partial y_i}{\partial a_j} = \left\{\begin{array}{ll} \dfrac{\exp(a_j)(1-y_i)}{\sum_j \exp(a_j)} = y_i(1-y_i), & \text{if } i = j, \\ -\dfrac{\exp(a_j) y_i}{\sum_j \exp(a_j)} = -y_i y_j, & \text{if } i \neq j \end{array}\right\} = y_i(\delta_{ij} - y_j).$$