Problem 1

(a) Ridge Regression

$$L(\vec{w}) = \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^2 + \lambda \left|\left|\vec{w}\right|\right|_2^2 = \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^T (\vec{w}^T \vec{x}_i - y_i) + \lambda \vec{w}^T \vec{w}$$

$$= \sum_{i=1}^{n} \left(y_i^T y_i - 2\vec{w}^T \vec{x}_i{}^T y_i + \vec{w}^T \vec{x}_i^T \vec{x}_i \vec{w} \right) + \vec{w}^T \lambda I \vec{w}$$

$$\frac{\partial L(\vec{w})}{\partial \vec{w}} = \sum_{i=1}^{n} \left(-2\vec{x}_i^T y_i + 2\vec{x}_i^T x_i \vec{w} \right) + 2\lambda I \vec{w}$$

$$\vec{w}_{j+1} = \vec{w}_j - c \cdot \frac{\partial L(\vec{w}_j)}{\partial \vec{w}_j} = \vec{w}_j - c \cdot \left(\sum_{i=1}^{n} \left(-2\vec{x}_i^T y_i + 2\vec{x}_i^T x_i \vec{w}_j \right) + 2\lambda I \vec{w}_j \right)$$

(b) Lasso Regression

$$L(\vec{w}) = \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^2 + \lambda |\vec{w}| = \sum_{i=1}^{n} (\vec{w}^T \vec{x}_i - y_i)^T (\vec{w}^T \vec{x}_i - y_i) + \lambda \sum_{\alpha=1}^{d} |w_\alpha|$$

$$= \sum_{i=1}^{n} \left(y_i^T y_i - 2\vec{w}^T \vec{x}_i{}^T y_i + \vec{w}^T \vec{x}_i^T \vec{x}_i \vec{w} \right) + \lambda \sum_{\alpha=1}^{d} |w_\alpha|$$

Because $L1$ regularizer $\sum_{\alpha=1}^{d} |w_\alpha|$ is not differentiable, the subgradient is

$$\frac{\partial L(\vec{w})}{\partial \vec{w}} = \sum_{i=1}^{n} \left(-2\vec{x}_i^T y_i + 2\vec{x}_i^T x_i \vec{w} + \lambda\, sign(\vec{w}) \right)$$

$$\vec{w}_{j+1} = \vec{w}_j - c \cdot \frac{\partial L(\vec{w}_j)}{\partial \vec{w}_j} = \vec{w}_j - c \cdot \left(\sum_{i=1}^{n} \left(-2\vec{x}_i^T y_i + 2\vec{x}_i^T x_i \vec{w}_j + \lambda\, sign(\vec{w}_j) \right) \right)$$

(c) Logistic Regression

$$L(\vec{w}) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \vec{w}^T \vec{x}_i))$$

$$\frac{\partial L(\vec{w})}{\partial \vec{w}} = \sum_{i=1}^{n} \frac{\exp(-y_i \vec{w}^T \vec{x}_i)'}{1 + \exp(-y_i \vec{w}^T \vec{x}_i)} = \sum_{i=1}^{n} \frac{-\vec{x}_i^{\mathrm{T}} \mathrm{y_i} \exp(-y_i \vec{w}^T \vec{x}_i)}{1 + \exp(-y_i \vec{w}^T \vec{x}_i)}$$

$$\vec{w}_{j+1} = \vec{w}_j - c \cdot \frac{\partial L(\vec{w}_j)}{\partial \vec{w}_j} = \vec{w}_j - c \cdot \left(\sum_{i=1}^{n} \frac{-\vec{x}_i^{\mathrm{T}} \mathrm{y_i} \exp(-y_i \vec{w}_j{}^T \vec{x}_i)}{1 + \exp(-y_i \vec{w}_j{}^T \vec{x}_i)} \right)$$

(d) Linear Support Vector Machine

$$L(\vec{w}) = C\sum_{i=1}^{n} \max\{1 - y_i \vec{w}^T \vec{x}_i, 0\} + \left||\vec{w}|\right|_2^2$$

For each piece of the training data

$$\frac{\partial l_i(\vec{w})}{\partial \vec{w}} = \begin{cases} 0, & y_i \vec{w}^T x_i \geq 1 \\ -y_i x_i, & y_i \vec{w}^T x_i < 1 \end{cases}$$

$$\frac{\partial L(\vec{w})}{\partial \vec{w}} = \sum_{i=1}^{n} \frac{\partial l_i(\vec{w})}{\partial \vec{w}}$$

$$\vec{w}_{j+1} = \vec{w}_j - c \cdot \frac{\partial L(\vec{w}_j)}{\partial \vec{w}_j} = \vec{w}_j - c \cdot \left(\sum_{i=1}^{n} \frac{\partial l_i(\vec{w})}{\partial \vec{w}}\right)$$