Problem 1

Let max be the element-wise maximum, $\lambda > 0$. Show that if

$$w^* = \min_{w \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} l(y_i, wx_i) + \frac{\lambda}{2} \|w\|_2^2$$

where l is convex, then $w^* \in \operatorname{span}(\{x_i\}_{i=1}^m)$. This means that for a high-dimensional data matrix of low rank, we can reduce the dimensionality of the data matrix without changing the corresponding weights.

Proof Suppose for the sake of contradiction that $w^* = \sum_{i=1}^n a_i x_i + u$, where a_i are constants and u is some nonzero vector orthogonal to $\operatorname{span}(\{x_i\}_{i=1}^m)$. Let $w = \sum_{i=1}^n a_i x_i$. By the orthogonality of u, (1) $w^* x_i = wx_i$ for all i, so $l(y_i, w^* x) = l(y_i, wx)$ and (2) $\|w^*\|_2^2 = \|w\|_2^2 + u^2 > \|w\|_2^2$, contradicting that w^* minimizes the regularized loss.

Problem 2

Let Π_K denote the projection operation onto the set K. We wish to show using Lagrange multipliers that if we have a constrained set K, and convex function f, then x^* minimizes f over K iff $\Pi_K [x^* - \nabla f(x^*)] = x^*$.

Proof Suppose K does not contain the unconstrained global minimum. Then for all $x \in K$, $\nabla f(x) \neq 0$. Then f must be monotonic along any line through K, so there must be some x^* on the boundary of K that minimizes f. Let $g(x) = c$ define the boundary of K.

First we show that if x^* minimizes f over K, then $\Pi_K(x^* - \lambda \nabla g(x^*)) = x^*$. We have by the Lagrangian multiplier condition that the solution x^* satisfies $\nabla f(x^*) = \lambda \nabla g(x^*)$. If x^* is the constrained minimum, observe that $\Pi_K(x^* - \nabla f(x^*)) = \Pi_K(x^* - \lambda \nabla g(x^*))$. It must be the case that $\lambda \leq 0$. Suppose for the sake of contradiction that $\lambda > 0$. By convexity $f(x) \geq f(x^*) + \nabla f(x^*)(x - x^*)$ for any $x \in K$. Since $f(x^*)$ is a minimum,

$$\nabla f(x^*)(x - x^*) \geq 0$$

We can show that $\nabla g(x^*)$ is a supporting hyperplane, so

$$\nabla g(x^*)(x - x^*) \leq 0$$

for any $x \in K$ But $\nabla f(x^*)(x - x^*) = \lambda \nabla g(x^*)(x - x^*) \leq 0$, and $\nabla f(x) \neq 0$, a contradiction. Thus $\lambda \leq 0$. Then $\Pi_K(x^* - \nabla f(x^*)) = \Pi_K(x^* - \lambda \nabla g(x^*)) = x^*$ by the orthogonality of $\nabla g(x^*)$.

Next we show the reverse implication also holds. If $\Pi_K [x^* - \nabla f(x^*)] = x^*$, x^* must be on the boundary of K and by the Pythagorean theorem $\nabla f(x^*) = \lambda \nabla g(x^*)$ for some $\lambda \leq 0$. Then the convexity of f and that $\nabla g(x)$ is a supporting hyperplane gives us for any $x \in K$

$$f(x) \geq f(x^*) + \nabla f(x^*)(x - x^*) = f(x^*) + \lambda \nabla g(x^*)(x - x^*) \geq f(x^*)$$

and we are done.

Suppose K contains the unconstrained global minimum x^*. Then we must have $\nabla f(x^*) = 0$, so $\Pi_K [x^* - \nabla f(x^*)] = x^*$. Conversely, if $\Pi_K [x^* - \nabla f(x^*)] = x^*$, then $x^* - \nabla f(x^*) \notin K$. Then x^* must be on the boundary of K by the Pythagorean theorem, and by the above arguments x^* is the minimum of f over K.

1 Show that $\nabla g(x^*)$ is a supporting hyperplane

Supporting Hyperplane Theorem For a nonempty convex set C, if x is on the boundary of C, then there exists a supporting hyperplane of C passing through x. In other words, there is a vector $a \in \mathbb{R}^n, a \neq 0$ such that $\sup_{z \in C} a^T z \leq a^T x$.

Proof For a sequence $\{x_k\} \not\subseteq C$ such that $x_k \to x$, where x is a point on the boundary of C, and a sequence of projections $\{z_k\}$ of x_k on the set C, $a_k = \frac{x_k - z_k}{\|x_k - z_k\|}$ converges to a supporting hyperplane. No points of the form $x^* + \epsilon \nabla g(x^*), \epsilon > 0$ are in C, as $\nabla g(x^*)(x^* + \epsilon \nabla g(x^*)) = \nabla g(x^*) x^* + \epsilon \|\nabla g(x^*)\|^2 \geq 0$. Then we can construct a sequence $\{x_k = x^* + \epsilon_k \nabla g(x^*)\}$ such that $x_k \to x^*$. Let z_k be the projection of x_k on the set C for each k. Clearly, $a = \frac{x_k - z_k}{\|x_k - z_k\|} = \frac{\nabla g(x^*)}{\|g(x^*)\|}$ for all k.

Citation: `http://www.ifp.illinois.edu/~angelia/L7_separationthms.pdf`
The proof is mostly from this link. I added the note about $a = \nabla g(x^*)$ satisfying the theorem (It is somewhat obvious, admittedly).