Problem 1

The Lagrangian function and its derivatives for β, β_0 and ξ_i are:

$$
\begin{aligned}
L &= \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i[y_i(x_i^T\beta+\beta_0)-(1-\xi_i)] - \sum_{i=1}^{N}\mu_i\xi_i \\
\nabla_\beta &= \beta - \sum_{i=1}^{N}\alpha_i y_i x_i \\
\nabla_{\beta_0} &= -\sum_{i=1}^{N}\alpha_i y_i \\
\nabla_{\xi_i} &= C - \alpha_i - \mu_i
\end{aligned}
$$

Set the derivatives to 0, we get $\sum_{i=1}^{N}\alpha_i y_i = 0,\ \beta = \sum_{i=1}^{N}\alpha_i y_i x_i,\ \alpha_i = C - \mu_i$.

$$
\begin{aligned}
L_D &= \frac{1}{2}\beta^T\beta + C\sum_{i=1}^{N}\xi_i - \beta^T\beta + \sum_{i=1}^{N}[\alpha_i(1-\xi_i)-\mu_i\xi_i] \\
&= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j
\end{aligned}
$$

The dual problem is $\max\limits_{\alpha} L_D, s.t.\ 0 \le \alpha_i \le C,\ \sum_{i=1}^{N}\alpha_i y_i = 0$.

Then we can get $\hat{\alpha}_i$ by optimizing the dual problem. The solution for the primal problem is:

$$
\hat{\beta} = \sum_{i=1}^{N}\hat{\alpha}_i y_i x_i,\ \hat{\beta}_0 = y_{i^*} - x_{i^*}\hat{\beta}
$$

where i^* is a margin point($\hat{\alpha}_{i^*} > 0, \hat{\xi_{i^*}} = 0$).

Problem 2

Let $f(\beta,\beta_0) = \sum_{i=1}^{n}(y-\beta_0-x^T\beta)^2,\ g(\beta) = \sum_{i=1}^{p}|\beta_j| - s$.

The primal problem is $\min\limits_{\beta_0,\beta} f(\beta,\beta_0),\ s.t.\ g(\beta) \le 0$ and the dual problem is $\max\limits_{\lambda}\theta(\lambda), \lambda \ge 0$, where $\theta(\lambda) = \inf\limits_{\beta_0,\beta}\{f(\beta,\beta_0)+\lambda g(\beta)\}$.

Noted that f and g satisfy the condition of the strong dual theorem, if $\bar{\beta},\bar{\beta}_0$ is a feasible solution of the primal problem, then there exists some $\bar{\lambda}$, s.t.

$$
\begin{aligned}
f(\bar{\beta},\bar{\beta}_0) &= \theta(\bar{\lambda}) = \sup\{\theta(\lambda), \lambda \ge 0\} \\
\bar{\lambda}g(\bar{\beta}) &= 0
\end{aligned}
$$

So $f(\bar{\beta}, \bar{\beta}_0) + \bar{\lambda} g(\bar{\beta}) = f(\bar{\beta}, \bar{\beta}_0) = \theta(\bar{\lambda})$. In other words, β_0, β is also a feasible solution of $\min_{\beta_0,\beta} f(\beta, \beta_0) + \lambda g(\beta)$.
Further more,

$$\underset{\beta_0,\beta}{argmin}\, f(\beta, \beta_0) + \lambda g(\beta) = \underset{\beta_0,\beta}{argmin} \sum_{i=1}^{n} (y - \beta_0 - x^T \beta)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

So the promal problem is equivalent to

$$\min_{\beta_0,\beta} \sum_{i=1}^{n} (y - \beta_0 - x^T \beta)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

for some $\lambda > 0$.

Problem 3

Please see attachment.