CECS 551 Exam 2, Spring 2016, Dr. Ebert

Category C

1. State the Lagrange dual (maximization problem) of the primal maximum-margin optimization problem with respect to training set $\mathcal{D} = \{(\overline{x}_1, y_1), \ldots, (\overline{x}_l, y_l)\}$. Remember to include the dual constraints.

2. How does your answer to Problem 1 change when slack variables and slack cost C are introduced into the primal problem?

3. Suppose $\overline{w}^*$ is the normal to the maximum-margin decision hyperplane, and $\overline{x}_{\mathrm{SV}+}$ is a positive support vector. Provide the equation for obtaining the offset term b^*.

4. After solving the Lagrange dual maximization problem for soft-margin classifiers, what does the Lagrange multiplier α_i say about the corresponding training vector $\overline{x}_i$ in relation to the optimal decision surface?

5. Given a training set of 500 vectors, provide the sizes of the training and testing sets for each fold of 10-fold cross validation.

6. The 90% confidence interval for the 10-fold-cross-validation error for some model has been determined as $[0, 0.13]$. If the model is deployed each day over a 30-day period, and the error rate is measured for each day, about how many days should we expect the daily error rate to equal or exceed 13%? Explain.

7. With respect to non-linear svm classifiers, what is the difference between "input space" and "feature space". Describe the "kernel trick" that is being played by a positive-definite kernel when it comes to classifying a vector $\overline{x}$ from the input space.

8. State the difference between the expected risk versus the empirical risk of a model $\hat{f}$ with respect to a data universe $\mathcal{X}$ and training set $\mathcal{D}$. In general, what is the only way to guarantee that the two risks are as close to each other as one desires, with high probability?

9. State one advantage and one disadvantage of i) one-versus-the-rest multiclass classification and ii) pairwise multiclass classification.

10. After solving the Lagrange dual maximization problem for the optimal soft-margin regression model, what do the Lagrange multipliers α_i and α'_i say about the corresponding training vector $\overline{x}_i$?

Category B

B1a. Assume $k_1 : \mathcal{R}^n \times \mathcal{R}^n \to \mathcal{R}$ and $k_2 : \mathcal{R}^n \times \mathcal{R}^n \to \mathcal{R}$ are both kernels, and $c_1, c_2 \geq 0$ are constants. Then prove that $c_1 k_1 + c_2 k_2$ is a kernel, where, e.g.,

$$(c_1 k_1 + c_2 k_2)(\overline{x}, \overline{y}) = c_1 k_1(\overline{x}, \overline{y}) + c_2 k_2(\overline{x}, \overline{y}).$$

B1b. Given constant $c \geq 0$, prove that the constant function $k(\overline{x}, \overline{y}) = c$ is a kernel.

B2a. Provide a formula for the solution $\overline{w}^*$ to the maximum-margin ϵ-regression optimization problem, in terms of the dual solution.

B2b. The formula for the solution b^* to the maximum-margin ϵ-regression optimization problem is

$$b^* = \frac{1}{l}\sum_{i=1}^{l}(\overline{w}^* \cdot \overline{x}_i - y_i).$$

This can be viewed as a special case of the following result: the value of x that minimizes

$$S(x) = \sum_{i=1}^{l}(x - a_i)^2$$

is

$$x = \frac{1}{l}\sum_{i=1}^{l} a_i,$$

where $a_1, \ldots, a_l$ are constants. Prove this result. Hint: $S(x)$ is a quadratic function.

B3a. If a data set consists of the positive vector $(4, 2)$, and the negative vector $(0, 0)$, then provide the equation (in the form $w_1 x + w_2 y = b$) for the maximum-margin decision line.

B3b. Solve the Lagrange dual of the maximum-margin optimization problem for the data set from part a. Use the solution to determine both $\overline{w}^*$ and b^*, and verify that these answers are consistent with your answer from part a.

Category A

A1. Given the primal minimization problem

$$\min_{\overline{x}} \phi(\overline{x}),$$

subject to

$$g_i(\overline{x}) \geq 0,$$

forall $i = 1, \ldots, l$, suppose that, for the optimal Lagrange dual solution $(\overline{\alpha}^*, \overline{x}^*)$, all the KKT conditions are satisfied. Moreover, suppose that there is a constraint g_j, $1 \leq j \leq l$, for which both $g_j(\overline{x}^*) = 0$, and $\alpha_j^* = 0$. Prove that the removal of constraint g_j from the primal problem does not change the primal optimal solution $\overline{x}^*$. Hint: use a proof by contradiction and a viewpoint that max and min are "adversaries".

A2a. Consider the problem of minimizing real-valued function $f(x)$ subject to $x \geq b$, for some real constant b. Assume the following about $f(x)$.

- $f'(x)$ exists and is equal to some function $g(x)$.
- $g(x)$ has an inverse $g^{-1}(x)$.
- g^{-1} is differentiable with derivative equal to some function $h(x)$.

Provide the Lagrangian $L(\alpha, x)$ of this optimization problem. Assuming that all KKT conditions are satisfied, show that the primal solution x^* satisfies $x^* = g^{-1}(\alpha^*)$.

A2b. Provide the Lagrange dual $\phi(\alpha)$, and show that it has a critical point at $\alpha = g(b)$. Assuming this critical point maximizes ϕ, show that $\overline{x}^* = b$.

A3. The Lagrangian of the maximum soft-margin classifier problem is

$$L(\overline{\alpha}, \overline{\beta}, \overline{w}, \overline{\xi}, b) = \frac{1}{2}\overline{w} \cdot \overline{w} + C\sum_{i=1}^{l} \xi_i$$

$$-\sum_{i=1}^{l} \alpha_i(y_i(\overline{w} \cdot \overline{x}_i - b) + \xi_i - 1) - \sum_{i=1}^{l} \beta_i \xi_i.$$

Prove that the objective function of the dual of this problem is identical to that of the dual of the hard-margin classifier problem.