

Written Assignment 4

This assignment is due by 11/20 11:59pm. Please submit solutions as a single PDF file through Canvas. Handwritten or typed solutions are fine, so long as the solution is legible and is submitted as a PDF. You might find LaTeX to be a useful tool for typesetting text and math expressions.

1. In this question we consider the hard margin SVM formulation as developed in class and the textbook. Recall that we stated that for this type of optimization problem there is no duality gap. That is, the primal and dual objective functions have the same value at their optimal solutions. Use this fact and other facts developed in class to show that the maximum margin γ^* can be calculated from the dual solution α^* . More specifically show that $\frac{1}{(\gamma^*)^2} = \sum_i \alpha_i^*$.
2. The soft margin SVM formulation developed in class called for minimizing $\frac{1}{2}V^TV + C \sum_i \xi_i$ subject to constraints (for all i) $1 - \xi_i - t_i(v^T \phi(x_i) + v_0) \leq 0$ and $\xi_i \geq 0$.

Consider the alternative formulation QSVM that uses a quadratic penalty on ξ_i which is: minimize $\frac{1}{2}V^TV + \frac{1}{2}C \sum_i \xi_i^2$ subject to constraints (for all i) $1 - \xi_i - t_i(v^T \phi(x_i) + v_0) \leq 0$. Note that we do not need the constraint $\xi_i \geq 0$ in this formulation because $\xi_i < 0$ increases the objective function and makes the constraints harder to satisfy. Develop the dual formulation of QSVM: that is (1) compute the dual objective function, (2) state the dual optimization problem, and (3) argue that QSVM is also a kernel method.

3. In this problem we show that many machine learning problems have solutions that can be expressed through kernels. Consider the standard formulation where $a_i = w^T \phi(x_i)$, and where we have a loss function $\ell(a_i, t_i)$. For example we might use the square loss $\ell(a_i, t_i) = (t_i - a_i)^2$ or the classification log loss $\ell(a_i, t_i) = t_i \ln y_i + (1 - t_i) \ln(1 - y_i)$ where $y_i = \sigma(a_i)$ etc. But for this question we consider any loss function. In addition consider any regularization function $R(b)$ which is monotonically increasing, where b is a scalar.

Now consider a machine learning objective of the form $\sum_i \ell(a_i, t_i) + R(w^T w)$.

Show the the optimal solution can be expressed as $w = \sum_i \alpha_i \phi(x_i)$ that is, it is a weighted sum of training examples. Note that this shows that predictions on test example z can be computed as $w^T \phi(z) = \sum_i \alpha_i K(x_i, z)$.

4. In lecture we provided a general form for optimizing the hyper-parameters of Gaussian Process Regression (GPR). In general there is no closed form and we must use some iterative gradient based optimization. Consider GPR where one hyperparameter, θ_0 , gives the overall scale of the covariance. Specifically, assume the covariance matrix has the form $C(\mathbf{x}_n, \mathbf{x}_m) = \theta_0 k(\mathbf{x}_n, \mathbf{x}_m)$ where $k(\cdot, \cdot)$ is any valid kernel. Show that the marginal likelihood can be optimized w.r.t. θ_0 in closed form and develop the solution for θ_0 .