CS181 Practice Questions: Max-Margin Classification

When maximizing the margin, we seek to learn linear functions of the form

$$f(\boldsymbol{x}, \boldsymbol{w}, b) = \boldsymbol{w}^{\mathsf{T}} \boldsymbol{\phi}(\boldsymbol{x}) + b$$

where $\boldsymbol{w}$ is an M-dimensional column vector of weights and $\boldsymbol{\phi}(\boldsymbol{x})$ is a collection of feature maps (like the regression and neural network case). The training data set comprise of N input vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ with the corresponding target labels $t_1, t_2, \ldots, t_N$, where $t_n \in \{-1, +1\}$.

1. **Computing the Margin**

   What is the perpendicular distance from a data point $\mathbf{x_n}$ to the decision boundary $y_w(x)$?

2. **Basic Maximization Problem**

What is the optimization problem we write for maximizing the margin?

3. **Constrained Minimization**

What is the corresponding constrained quadratic minimization problem for maximizing the margin?

4. **Equivalence**

Explain (at a high level) why the constrained quadratic minimization of question (3) is equivalent to the unconstrained maximization in (2)

5. **Tightness**

What happens to the inequalities $t_n(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_n) + b) \geq 1$ for the optimal solution?

6. **Kernels**

What is kernel function?