第三次作业

一、(线性回归) 给定高斯先验分布 $p(\beta)=N(0,\tau I)$ 和高斯采样模型 $p(y|x,\beta)=N(x^T\beta,\sigma^2 I)$，利用贝叶斯公式计算后验概率 $p(\beta|y,x)$，并找出岭回归公式中正则化参数 λ 与方差 τ 和 σ^2 的关系。

二、(核方法) 设 $\{\phi_i(x)\}_{i=1}^{\infty}$ 是 Hilbert 函数空间 H 的完备正交基，$K(x,y)=\sum_{i=1}^{\infty}\gamma_i\phi_i(x)\phi_i(y)$ 是一个核函数。对于任意 H 空间的两个函数 $f(x)=\sum_{i=1}^{\infty}a_i\phi_i(x)$ 和 $g(x)=\sum_{i=1}^{\infty}b_i\phi_i(x)$，定义核再生空间 H_K 的内积为 $<f,g>_{\mathrm{H}_K}=\sum_i a_i b_i/\gamma_i$。证明

$$<K(.,x_i),f>_{\mathrm{H}_K}=f(x_i)\,;\qquad <K(.,x_i),K(.,x_j)>_{\mathrm{H}_K}=K(x_i,x_j)\,.$$

三、(核方法) 设训练数据 $\{(x_i,y_i)\}_{i=1}^{L}$，其中 $x_i\in R^p, y_i\in R^1$. 考虑在 x_0 点的局部线性回归

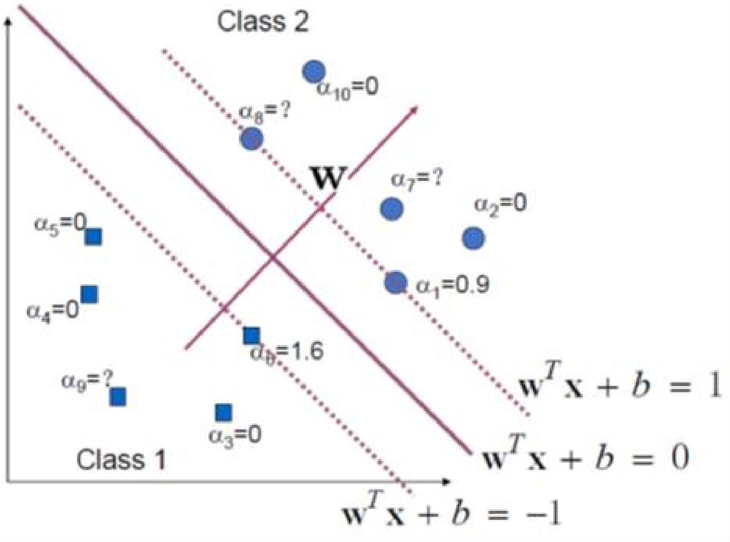
$$\min_{\beta}\sum_{i=1}^{N}K_\lambda(x_0,x_i)\left[y_i-[1\ x_i^T]\beta\right]^2$$

1. 给出回归参数 β 的解表达式。
2. 设回归函数 $\hat{f}(x_0)=\left[1\ x_0^T\right]\hat{\beta}(x_0)=\sum_{i=1}^{N}l_i(x_0)y_i$，证明

$$\sum_{i=1}^{N}l_i(x_0)=1\,;\qquad \sum_{i=1}^{N}x_i l_i(x_0)=x_0$$

四、(支持向量) 设训练数据 $\{(x_i,y_i)\}_{i=1}^{L}$，其中 $x_i\in R^n, y_i\in\{-1,1\}$，是线性可分的。

1. 建立一个分类器使得训练数据到分界面的间隔极大优化问题的数学表述；
2. 利用 Lagrange 乘子法将给问题转化为二次规划问题，给出支持向量的几何解释；
3. 下图所示的参数 $\{\alpha_i\}_{i=1}^{10}$ 是 Lagrange 乘子，填充缺失的 $\alpha_7,\alpha_8,\alpha_9$ 的值。



五、（模型选择）： 假设统计模型 $Y=f(X)+\varepsilon$，其中误差ε服从正态分布$N(0,\sigma_{\varepsilon})$，$\hat{f}(X)$试相应的回归拟合函数。对于任意输入点$X=x_0$，给出其期望预测误差$Err(x_0)=E\left[(Y-\hat{f}(x_0))^2 \mid X=x_0\right]$的分解表达式（分解成为偏倚和方差）。进一步证明在平方误差损失情况下，训练误差的乐观性有下列估计

$$\omega = E_y(op)=E_y\left[Err_{in}-\overline{err}\right]=\frac{2}{N}\sum_{i=1}^{N}Cov(\hat{y}_i,y_i)$$