Practice Midterm 2010

4.) One-class SVM

a)

$$\min_{w} \frac{1}{2} w^T w$$

s.t. $w^T x^{(i)} \geq 1$

$\Rightarrow g(i) = -w^T x^{(i)} + 1 \leq 0$

$$\Rightarrow L(w,\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{m} \alpha_i \left[w^T x^{(i)} - 1 \right] \qquad \text{I.}$$

$$\Rightarrow \nabla_w L(w,\alpha) = w - \sum_{i=1}^{m} \alpha_i x^{(i)} = 0$$

$$\Rightarrow w = \sum_{i=1}^{m} \alpha_i x^{(i)} \qquad \text{II.}$$

$\Rightarrow$ II. in I:

$$L(w,\alpha) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j x^{(i)T} x^{(j)} - \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j (x^{(i)})^T x^{(j)} + \sum_{i=1}^{m} \alpha_i$$

$$\Rightarrow L(w,\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j x^{(i)T} x^{(j)}$$

$$\Rightarrow \boxed{\text{Dual opt problem: } \max_{\alpha} W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle}$$

s.t. $\alpha_i \geq 0, \; i = 1, \dots, m$

Practice Midterm 2010

4b) Can the one-class SVM be Kernelized?

Yes, because we only use inner product to train and test

c) SMO can iterate over all α_i's independently because all convex optimization constraints are independent.

$\Rightarrow$ Maximize $W(\alpha)$ wrt α_i:

$$\nabla_{\alpha_i} W(\alpha) = 1 - \frac{1}{2} \sum_{\substack{j=1 \\ i \neq j}}^{m} \alpha_j \langle x^{(i)}, x^{(j)} \rangle - \frac{1}{2} \cdot 2 \cdot \alpha_i \langle x^{(i)}, x^{(i)} \rangle = 0$$

$$\Rightarrow 1 = \frac{1}{2} \sum_{\substack{j=1 \\ i \neq j}}^{m} \alpha_j \langle x^{(i)}, x^{(j)} \rangle + \alpha_i \langle x^{(i)}, x^{(i)} \rangle$$

$$\alpha_i = \frac{1 - \frac{1}{2} \sum_{\substack{j=1 \\ i \neq j}}^{m} \alpha_j \langle x^{(i)}, x^{(j)} \rangle}{\langle x^{(i)}, x^{(i)} \rangle}$$

229 Practice Midterm 2010

2.)

$$\theta_{MAP} = \arg\max_{\theta} \left(\prod_{i=1}^{m} p(y^{(i)} | x^{(i)}, \theta) \right) \cdot p(\theta)$$

$$p(y^{(i)} | x^{(i)}, \theta) = \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(- \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\theta^2} \right)$$

$$p(\theta_i) = \frac{1}{\sqrt{2\pi}\,\tau_i} \exp\left(\frac{-\theta_i^2}{2\tau^2} \right)$$

$$\Rightarrow L(\theta) = \left(\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(- \frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\theta^2} \right) \right) \prod_j \frac{1}{\sqrt{2\pi}\,\tau} \cdot \exp\left(\frac{-\theta_j}{2\tau^2} \right)$$

$$\Rightarrow \log(L(\theta)) = \ell(\theta) = m \cdot \log \frac{1}{\sqrt{2\pi}\,\theta} + n \cdot \log \frac{1}{\sqrt{2\pi}\,\tau}$$

$$- \frac{1}{\theta^2} \cdot \frac{1}{2} \sum_{i=1}^{m} (y^{(i)} - \theta^T x^{(i)})^2$$

$$- \frac{1}{\tau^2} \cdot \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

Practice Midterm 2010

$\Rightarrow$ To maximize $\ell(\theta)$, need to minimize:

$$\frac{1}{\sigma^2}\frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)}-\theta^{(T)}x^{(i)}\right)^2+\frac{1}{\tau^2}\frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

~~can ignore $\frac{1}{\sigma^2}$ and $\frac{1}{\tau^2}$ because they are fixed $\Rightarrow$ need to minimize vector~~

$$\Rightarrow \nabla_\theta \ell(\theta)=\frac{1}{\sigma^2}\nabla_\theta \frac{1}{2}(X\theta-\vec{y})^T(X\theta-\vec{y})+\frac{1}{\tau^2}\nabla_\theta\frac{1}{2}\theta^T\theta$$

$$=\frac{1}{\sigma^2}\left(X^TX\theta-X^T\vec{y}\right)+\frac{1}{\tau^2}\cdot\vec{\theta}$$

$$\cancel{\frac{1}{\tau^2}\cdot\vec{\theta}}$$

Take:

$\frac{1}{\sigma^2}\cdot I=S$

$\frac{1}{\tau^2}\cdot I=J$

$$\Rightarrow \nabla_\theta \ell(\theta)=S\cdot X^TX\cdot\theta-S\cdot X^T\cdot\vec{y}+J\cdot\theta=0$$

$$\Rightarrow S\cdot X^T\cdot X\cdot\theta+J\cdot\theta=S\cdot X^T\cdot\vec{y}$$

$$\Rightarrow S\cdot X^T\cdot X+J=S\cdot X^T\cdot y\cdot\theta^{-1}$$

$$y^{-1}X+S^{-1}\cdot(X^T)^{-1}\cdot y^{-1}\cdot J=\theta^{-1}$$

$$\Rightarrow \cancel{\boxed{\theta=X^{-1}\cdot y+J^{-1}\cdot y\cdot X^T\cdot S}}$$

$$\boxed{\theta=X^{-1}y+J^{-1}\cdot S\cdot X^T\cdot y}$$