Madison Chamberlain

Ch9 Q1

<u>Without the kernel trick:</u>

Given a polynomial function: $y = \omega^{\mathbf{T}} + \Phi(x)$ where we need to find $\Phi(x)$.

*minimize cost function w/ respect to weights, $\omega^{*} = \arg(\min(\omega))\, J(\omega)$

$$\omega=(x^{T}x)^{-1}x^{T}y$$

$$\omega=(\Phi^{T}\Phi)^{-1}\Phi^{T}y$$

$$\omega=(\Phi^{T}\Phi + \lambda l)^{-1}\Phi^{T}y$$

*now to find the weights, we need to know Φ^{T} or $\Phi^{T}\Phi$

*both values are extremely hard to compute

<u>With the kernel trick:</u>

$$J(\omega) = \sum(y_n - \omega^{T}\Phi(x_n))^{2}+(\lambda/2)\sum||\omega||^{2}$$

$$\omega^{*}=(1/\lambda)\sum(y_n - \omega^{T}\Phi(x_n))\Phi(x_n)$$

* call $a_n = (1/\lambda)\sum(y_n - \omega^{T}\Phi(x_n))$

$$\omega^{*}=\sum(a_n\Phi(x_n))$$

$$\omega^{*}=\Phi^{T}a$$

*given the cost function: $J(\omega) = \sum(y_n - \omega^{T}\Phi(x_n))^{2}+(\lambda/2)\sum||\omega||^{2}$

$$J(\omega) = (y-\Phi\omega)^{T}(y-\Phi\omega) + (\lambda/2)\omega^{T}\omega$$

$$J(\omega) = y^{T}y - y^{T}\Phi\omega - \omega^{T}\Phi^{T}y + \omega^{T}\Phi^{T}\Phi\omega + (\lambda/2)\omega^{T}\omega$$

$$a= y^{T}y - y^{T}\Phi(\Phi^{T}a) - (\Phi^{T}a)^{T}\Phi^{T}y + (\Phi^{T}a)\Phi^{T}\Phi(\Phi^{T}a) + (\lambda/2)(\Phi^{T}a)^{T}(\Phi^{T}a)$$

$$J(a) = y^{T}y - y^{T}\Phi\Phi^{T}a - a^{T}\Phi\Phi^{T}y - a^{T}\Phi\Phi^{T}\Phi\Phi^{T}a + (\lambda/2)a^{T}\Phi\Phi^{T}a$$

*given that $\Phi\Phi^{T}$ = the kernel matrix K, where $K = K^{T}$ and $a^{T}Ka \geq 0$

$$J(a) = y^{T}y - y^{T}Ka - a^{T}Ky + a^{T}K^{2}a + (\lambda/2)a^{T}Ka$$

$$J(a) = y^{T}y - y^{T}Ka - y^{T}Ka + a^{T}K^{2}a + (\lambda/2)a^{T}Ka$$

*since $K = K^{T}$: $J(a) = y^{T}y - 2y^{T}Ka + a^{T}K^{2}a + (\lambda/2)a^{T}Ka$

*minimize:

$$(\delta(J(a))/(\delta(a) = 0 - 2y^{T}K + 2a^{T}K^{2} + \lambda a^{T}K = 0$$

$$-2y^{T} + 2a^{T}K + \lambda a^{T} = 0$$

$$a^{T}K + (\lambda/2)a^{T} = y^{T}$$

$$a^{T} = y^{T}(K + \lambda/2l)^{-1}$$

$$a^{*\,=}(K + \lambda/2l)^{-1}y$$

$$\omega= \Phi^{T}a = \Phi^{T}(K^{T}+\lambda^{1}l)^{-1}y = \Phi^{T}(\Phi\Phi^{T}+\lambda^{1}l)^{-1}y$$

*according to mercer's theorem, we can rewrite $\Phi\Phi^{T}$ as a function of its base features

This means we no longer need to be able to solve for Φ or $\Phi\Phi^{T}$