Din

(1)

For margin errors,

$$\xi_i^* = 1 - y_i ({w^*}^T x_i + b^*)$$

$\therefore$

$$y_i ({w^*}^T x_i + b^*) = 1 - \xi_i \quad (1)$$

$$w^T x_i = = ((1 - \xi_i)/y_i - b)$$

The eq. of hyperplane is

$$y_i (w^T x + b) = 1 \quad (2)$$

$$\text{ie } w^T x + b - 1/y_i = 0$$

~~Both of these~~ hyperplanes We know distance of a point (z) from a hyperplane is $(w^T x + b)$ is given by

$$\|r\| = \frac{|w^T z + b^*|}{\|w\|}$$

Here z is x_i

$$\therefore w^T x_i = \frac{1 - \xi_i}{y_i} - b$$

$$\& \quad b^* = b - 1/y_i$$

$$\therefore |r| = \frac{|(\cancel{1} - \xi_i)/y_i - \cancel{b} + \cancel{b} - \cancel{1/y_i}|}{\|w\|}$$

$$|r| = \frac{|\xi_i / y_i|}{\|w\|}$$

$$\therefore |r| \propto \xi_i$$

& The constant of proportionality

$$= \frac{1}{|y_i| \, \|w\|}$$

Q2

$\sigma = 4$

$C = 512$

CV_error = 0.2480

Test-error = 0.182

Support vector = 342.

Q3 Using Rayleigh Quotient

~~$\|X - PX\|_F^2 = Tr(X^TX) - Tr(X^TPX)$~~

~~where $P = AA^T$~~

$$\min_{\mu, A, \{\theta_i\}} \sum \|x_i - \mu - A\theta_i\|^2 = \min_{P \in P_K} \|X - PX\|_F \quad (1)$$

Now using rayleigh quotient

$$\|X - PX\|_F^2 = Tr((X-PX)^T(X-PX))$$

$$= Tr(X^TX) - Tr(X^TPX)$$

$$Tr(X^TPX) = \underbrace{Tr(A^TXX^TA)}$$

we need to maximize $Tr(A^TXX^TA)$ in order to minimize $\|X - PX\|_F$.

$\max_{A \in A_K} Tr(A^TCA)$ is

$A = [u_1 \ldots u_K]$

where $C = U\Lambda U^T$

$\therefore \max_{A \in A_k} tr(A^T X X^T A)$ is when

$A = [u_1 \; \ldots \; u_k]$

& $X X^T = \underbrace{U \Lambda U^T}_{d\times n \;\; n\times d}$

$\therefore tr(A^T X X^T A) = tr([u_1 \ldots u_k]^T U \Lambda U^T [u_1 \ldots u_k])$

$= tr[[u_1 \ldots u_k][u_1 \ldots u_k]^T U \Lambda U^T]$

$\cancel{[u_1 \ldots u_k]^T}$

$$\cancel{[u_1 \ldots u_k][u_1 \ldots u_k]^T = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & & 0 \end{pmatrix}}$$

$$\underbrace{[\underset{d\times 1}{u_1} \ldots u_k]^T}_{k\times d} \underbrace{[u_1 \ldots u_d]}_{d\times d} = \begin{bmatrix} 1 & & & & \\ & 1 & & 0 & \\ 0 & & 1 & & \\ & & & 1 \; 0 \; 0 & \end{bmatrix}_{k\times d}$$

k^{th} position.

$\parallel$ly.

$$\underbrace{[u_1 \ldots u_d]^T}_{d\times d} \underbrace{[u_1 \ldots u_k]}_{d\times k} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ 0 & 0 & 0 \\ \vdots & & \vdots \\ 0 & 0 & 0 \end{bmatrix}_{d\times k}$$

$$= tr\left(\begin{bmatrix} 1 & & & 0 \cdots 0 \\ & \ddots & & 0 \cdots 0 \\ & & 1 & 0 \cdots 0 \end{bmatrix}_{k \times d} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}_{d \times d} \begin{bmatrix} 1 & & 0 \\ 0 & \ddots & 1 \\ 0 & & 0 \end{bmatrix}_{d \times k} \right)$$

$$= n \sum_{i=1}^{k} \lambda_i$$

llly

$$tr(XX^T) = tr(U \Lambda U^T)$$

$$= tr(\underbrace{U^T U}_{nI} \Lambda)$$

$$= tr_d(n \Lambda)$$

$$= n \sum_{i=1}^{d} \lambda_i$$

$$\therefore \min_{\mu, A, \theta_i} \| x_i - \mu - A\theta_i \|^2 = n \sum_{i=1}^{d} \lambda_i - n \sum_{i=1}^{k} \lambda_i$$

$$= n \sum_{k+1}^{d} \lambda_i$$

b) ~~For the subspace A to be unique~~ ~~the solution to~~

~~We know, the~~ for subspace A to be unique the solution of above equation should be unique ie the eigenvalue of λ_k should be different from λ_{k+1}.

Other way to look at this is that

the solution to

$$\min \sum \|x_i - \mu - A\theta_i\|^2 = Tr(XX^T) - tr(A^TXX^TA)$$

Here $tr(XX^T)$ is deterministic

$\max tr(A^TXX^TA)$ is when A is $[u_1 \cdots u_k]$

When $\lambda_k = \lambda_{k+1}$ then there can be two possible choices for u_k & A will not be unique.

$\therefore$ Both necessary & sufficient cond is $\lambda_k \neq \lambda_{k+1}$ for A to be unique.

Q4 a)

	Principal components	% reduction
95% :	43	0.9787
99% :	167	0.9187

b) The 2nd principal component is capturing the case in which the lighting is lower left side of cheek. The 10th principal component is capturing nose, 18th is capturing people with big lips. So, the components are capturing different lighting conditions & features.