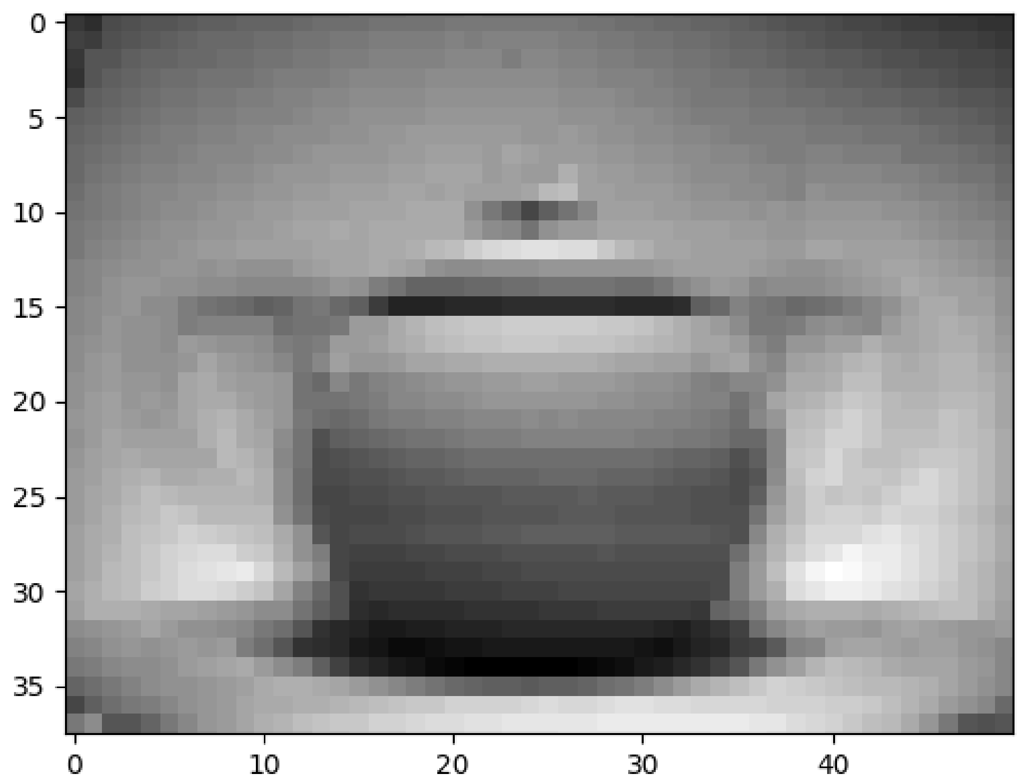
ML Assignment 4

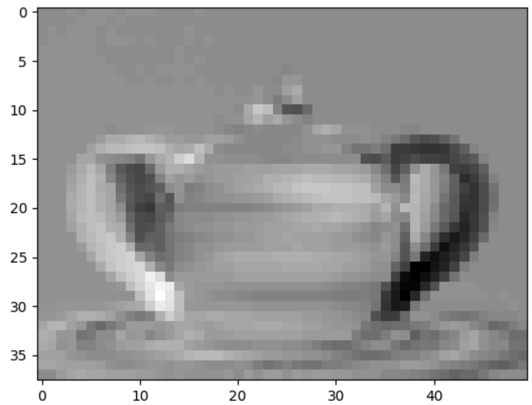
Simran Kucheria (sk11645@nyu.edu)

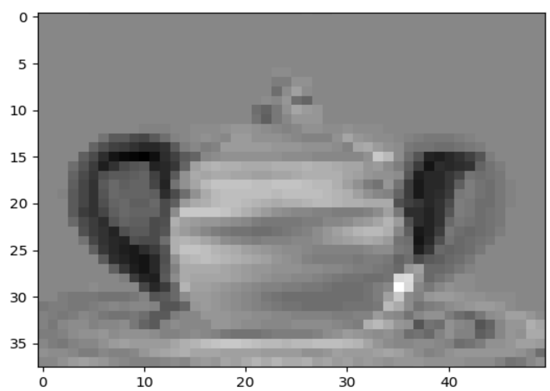
Problem 1

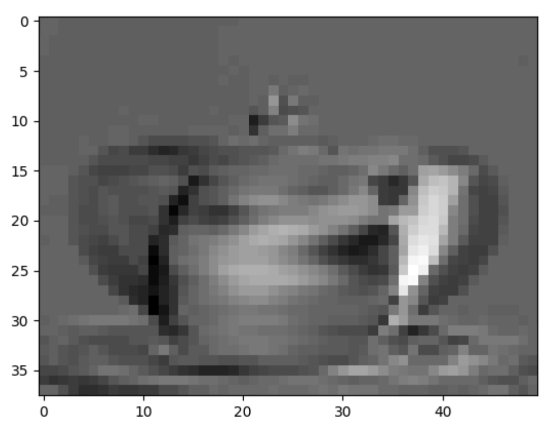
In order to apply PCA to the image dataset we first calculate the mean. The mean is represented as an image below.



Then we calculate the covariance matrix to further calculate eigenvectors and eigenvalues. The top 3 eigenvectors with eigenvalues of **4.215, 3.016, 2.099** are:







Using these eigenvectors and mean. We can reconstruct the teapot images.
$x_i \approx \mu + c_{ij} * v_j$ { mean + Coeff . Eigen Vectors)
Where $c_{ij} = (x_i - \mu)v_j$ { Centred Data . Eigen Vectors)

Used 10 random samples to show the comparison between the original and reconstructed images using the top 3 eigenvectors.

The performance of the reconstruction can be evaluated by calculating the least squares error which comes out to 13.6262. This can be reduced by increasing the number of eigenvectors inorder to encode more information.

Problem 2 and Problem 3

Q2. Box I → 8 apple 4 oranges

Box II → 10 apple 2 oranges.

$$\text{Prob}(\text{Box I} \mid \text{apple}) = \frac{\text{prob}(\text{apple} \mid \text{box I}) \cdot \text{prob}(\text{box I})}{\text{prob}(\text{apple})}$$

$$p(\text{box I}) = \frac{1}{2}$$

$$p(\text{apple} \mid \text{box I}) = \frac{8}{8+4} = \frac{8}{12}.$$

$$p(\text{apple}) = \frac{1}{2}\left(\frac{8}{12}\right) + \frac{1}{2}\left(\frac{10}{12}\right) = \frac{9}{12}.$$

$$P(\text{Box I} \mid \text{apple}) = \frac{\frac{8}{12} \cdot \frac{1}{2}}{9/12} = \frac{4}{9}.$$

Q3.

$$\theta = \{\alpha, u_1, \Sigma_1, u_2, \Sigma_2\}.$$

$$p(y \mid \theta) = \alpha^{y}(1-\alpha)^{1-y}.$$

$$p(x \mid y, \theta) = \mathcal{N}(x \mid u_y, \Sigma_y)$$

For an iid with N samples:

Joint probability $p(x, y \mid \theta) = \prod_{i=1}^{N} p(x_i \mid y_i, \theta) \cdot p(y_i \mid \theta)$

Taking log likelihood.

$$l(\theta) = \sum_{i=1}^{N} \log(p(y_i|\theta)) + \sum_{i=1}^{N} \log(p(x_i|y_i,\theta)).$$

$$= \sum_{i=1}^{N} \log(p(y_i|\theta)) + \sum_{y_i \in 1} \log(p(x_i|\mu_1,\Sigma_1)) + \sum_{y \in 2} \log(p(x_i|\mu_2,\Sigma_2))$$

MLE for

I α. $\rightarrow$ Take derivative w.r.t α and equate to 0

$$\frac{d(l(\theta))}{d\alpha} = \frac{d}{d\alpha}\left(\sum_{i=1}^{N} \log(p(y_i|\theta))\right)$$ $\rightarrow$ All other terms have no α term, hence derivative is 0 for them.

$$\log p(y_i|\theta) = \alpha^{y_i}(1-\alpha)^{(1-y_i)}.$$

$$\frac{d}{d\alpha}(l(\theta)) = \frac{d}{d\alpha}\left(\sum_{i=1}^{N}\left[\log(\alpha^{y_i}(1-\alpha)^{(1-y_i)})\right]\right).$$

$$= \frac{d}{d\alpha}\left(\sum_{i=1}^{N}\left[y_i \log\alpha + (1-y_i)\log(1-\alpha)\right]\right).$$

$$= \sum_{i=1}^{N}\left[\frac{y_i}{\alpha} + (-)\frac{(1-y_i)}{(1-\alpha)}\right].$$

$$= \sum_{i=1}^{N}\left(\frac{y_i - y_i\alpha - \alpha + y_i\alpha}{(\alpha)(1-\alpha)}\right)$$

$$= \sum_{i=1}^{N}\left(\frac{y_i - \alpha}{\alpha(1-\alpha)}\right) = 0 \qquad \sum_{i=1}^{N} y_i - N\alpha = 0$$

Let $\sum_{i=1}^{N} y_i$ be samples belonging N_1

$$N\alpha = \sum_{i=1}^{N} y_i$$

$$\alpha = \frac{N_1}{N}$$

II MLE for u_ϕ, u_2

Take derivative for u_0 & set to 0. Similarly do the same for u_1.

$$\frac{d\,(L(\theta)}{d u_\phi} = \frac{d}{du_0}\left[\sum_{y_i \in \phi} \log(p(x_i | u_\phi, \Sigma_\phi))\right]$$ (derivative of all other terms is 0 as no u_0 term).

$$\log(p(x_i | u_\phi, \Sigma_\phi)) = \log\left(\frac{1}{2\pi^{D/2}\sqrt{|\Sigma_\phi|}} e^{-1/2(\vec{x}_i - \vec{u}_\phi)^T \Sigma_\phi^{-1}(\vec{x}_i - \vec{u}_\phi)}\right)$$

$$-\frac{D}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_\phi|) - \frac{1}{2}(\vec{x}_i - \vec{u}_\phi)^T \Sigma_\phi^{-1}(\vec{x}_i - \vec{u}_\phi).$$

$$\frac{d(L(\theta)}{du_\phi} = \frac{d}{du_\phi}\left[\sum_{y_i=\phi_1}^{\phi}\left(\frac{-D}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_\phi|) - \frac{1}{2}(\vec{x}_i - \vec{u}_\phi)^T \Sigma_\phi^{-1}(\vec{x}_i - \vec{u}_\phi)\right)\right].$$

$$= \frac{d}{du_\phi}\left[\sum_{y_i \in \phi_1}\left(\frac{-1}{2}(\vec{x}_i - \vec{u}_\phi)^T \Sigma_0^{-1}(\vec{x}_i - \vec{u}_\phi)\right)\right]$$

$$\frac{d}{d\theta}(\theta^T\theta) = 2\theta^T \Rightarrow \text{property}$$

Applying property to eqn above

$$\frac{d(L(\theta)}{du_\phi} = \sum_{y_i \in \phi_1}\left(\frac{\cancel{-1}}{\cancel{2}}(\cancel{2}(\vec{x}_i - \vec{u}_\phi)^T \Sigma_\phi^{-1}\right) = \sum_{y_i \in \phi}(\vec{x}_i - \vec{u}_\phi)^T \Sigma_\phi^{-1}$$

Rewriting using trace notation:

$$\sum_{y_i \in \Phi} \left[\frac{-D}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_\Phi|) - \frac{1}{2} tr\left[(\vec{x_i} - \vec{u_\Phi})^T \Sigma_\Phi^{-1} (\vec{x_i} - \vec{u_\Phi}) \right] \right]$$

$$= \sum_{y_i \in \Phi} \left[\frac{-D}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma_\Phi|) - \frac{1}{2} tr\left[(\vec{x_i} - \vec{u_\Phi})^T (\vec{x_i} - \vec{u_\Phi}) \Sigma_\Phi^{-1} \right] \right]$$

$$= \sum_{y_i \in \Phi} \left[\frac{-D}{2} \log(2\pi) + \frac{1}{2} \log(|\Sigma_\Phi^{-1}|) - \frac{1}{2} tr\left[(\vec{x_i} - \vec{u_\Phi})^T (\vec{x_i} - \vec{u_\Phi}) \Sigma_\Phi^{-1} \right] \right]$$

Let $A = \Sigma_\Phi^{-1}$

$$= \sum_{y_i \in \Phi} \left[-\frac{D}{2} \log(2\pi) + \frac{1}{2} \log(|A|) - \frac{1}{2} tr\left[(\vec{x_i} - \vec{u_\Phi})^T (\vec{x_i} - \vec{u_\Phi}) A \right] \right]$$

$$\frac{d}{dA} \log|A| = (A^{-1})^T \qquad \frac{d}{dA} tr[BA] = B^T$$

Taking derivative wrt A and using the 2 properties

$$\frac{d(L(\sigma))}{dA} = \frac{d}{dA} \left[\sum_{y_i \in \Phi} \left(\frac{-D}{2} \log(2\pi) + \frac{1}{2} \log(|A|) - \frac{1}{2} tr\left[(\vec{x_i} - \vec{u_\Phi})^T (\vec{x_i} - \vec{u_\Phi}) A \right] \right) \right]$$

$$= \sum_{y \in \Phi} \frac{(A^{-1})^T}{2} - \frac{1}{2} \sum_{y \in \Phi} (\vec{x_i} - \vec{u_\Phi})^T (\vec{x_i} - \vec{u_\Phi})$$

$$= \sum_{y \in \Phi} \Sigma_\Phi - \sum_{y \in \Phi} (\vec{x_i} - \vec{u_\Phi})^T (\vec{x_i} - \vec{u_\Phi})$$

$$N_\Phi \Sigma_\Phi = \sum_{y \in \Phi} (x_i - u_\Phi)^T (\vec{x_i} - \vec{u_\Phi})$$

$$\Sigma_1 = \frac{1}{N_1}\left(\sum_{y\in 1} (\vec{x_i}-\vec{u_1})^T (\vec{x_i}-\vec{u_1})\right)$$

Similarly for Σ_1

$$\frac{d\,(l(\theta))}{d\Sigma_2} = \frac{d}{d\Sigma_2}\left[\sum_{y\in 2}\left(\frac{-D}{2}\log(2\pi)\right) - \frac{1}{2}\log(|\Sigma_2|) - \frac{1}{2}(\vec{x_i}-\vec{u_2})^T \Sigma_2^{-1}(\vec{x_i}-\vec{u_2})\right]$$

Rewriting using trace notation and replacing $\Sigma_1^{-1} = A$.

$$\frac{d\,(l(\theta))}{dA} = \frac{d}{dA}\left[\sum_{y\in 2}\left(\frac{-D}{2}\log(2\pi)\right) + \frac{1}{2}\log(|A|) - \frac{1}{2}\operatorname{tr}\left[(x_i-u_2)^T(x_i-u_2)A\right]\right]$$

$$= \sum_{y\in 2}\frac{(A^{-1})^T}{2} - \sum_{y\in 2}\frac{(\vec{x_i}-\vec{u_2})^T(\vec{x_i}-\vec{u_2})}{2} = 0.$$

$$N_2\Sigma_2 = \sum_{y\in 2}(\vec{x_i}-\vec{u_2})^T(\vec{x_i}-\vec{u_2})$$

$$\Sigma_2 = \frac{1}{N_2}\left(\sum_{y\in 2}(\vec{x_i}-\vec{u_2})^T(\vec{x_i}-\vec{u_2})\right)$$

For decision boundary

$$p(y=2|x) = p(y=1|x)$$

using bayes theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$P(y=2|x) = P(y=1|x)$$

$$\frac{P(x|y=2) \cdot P(y=2)}{\cancel{P(x)}} = \frac{P(x|y=1) \cdot P(y=1)}{\cancel{P(x)}}$$

$$P(x|y=2) \cdot P(y=2) = P(x|y=1) \cdot P(y=1)$$

$$P(x_i|y=2) = \mathcal{N}(x_i | u_2, \Sigma_2) \qquad p(y=2) = \alpha$$

$$p(x|y=1) = \mathcal{N}(x_i | u_\phi, \Sigma_\phi) \qquad p(y=1) = (1-\alpha)$$

$$\mathcal{N}(x_i | u_2, \Sigma_2) \cdot \alpha = \mathcal{N}(x_i | u_\phi, \Sigma_\phi) \cdot (1-\alpha)$$

Taking log on both sides

$$\log \alpha + \left[\frac{-D \log(2\pi)}{2} - \frac{1}{2} \log|\Sigma_1| - \frac{1}{2}(\vec{x_i} - \vec{u_2})^T \Sigma_1^{-1} (\vec{x_i} - \vec{u_2})\right] = \log(1-\alpha) + \left[\frac{-D\log(2\pi)}{2} - \frac{1}{2}\log|\Sigma_\phi| - \frac{1}{2}(\vec{x_i} - \vec{u_\phi})\Sigma_\phi^{-1} (\vec{x_i} - \vec{u_\phi})\right]$$

$$\log\left(\frac{\alpha}{1-\alpha}\right) + \frac{1}{2}\log\left|\frac{\Sigma_2}{\Sigma_1}\right| - \frac{1}{2}\left[(\vec{x_i} - \vec{u_2})^T \Sigma_2^{-1} (\vec{x_i} - \vec{u_2}) - (\vec{x_i} - \vec{u_\phi})^T \Sigma_\phi^{-1} (\vec{x_i} - \vec{u_\phi})\right] = 0$$

$$(\vec{x_i} - \vec{u_1})^T \Sigma_1^{-1} (\vec{x_i} - \vec{u_1}) = x_i^T x_i \Sigma_1^{-1} - 2 x_i \Sigma_1^{-1} u_1 + u_1^T \Sigma_1^{-1} \vec{u_1}$$

$$(\vec{x_i}-\vec{u_2})^T \Sigma_2^{-1} (\vec{x_i}-\vec{u_2}) - (\vec{x_i}-\vec{u_\phi}) \Sigma_\phi^{-1} (\vec{x_i}-\vec{u_\phi})$$

$$= x_i^T x_i \Sigma_2^{-1} - 2 x_i u_2 \Sigma_2^{-1} + u_2^T u_2 \Sigma_1^{-1} - (x_i \times \Sigma_{\phi}^{-1} - 2x_i u_\phi \Sigma_\phi^{-1} + u_\phi^T u_\phi \Sigma_\phi^{-1}$$

$$= x_i^T x_i (\underbrace{\Sigma_2^{-1} - \Sigma_\phi^{-1}}_{\to 0}) - 2x_i (u_2 \Sigma_2^{-1} - u_\phi \Sigma_\phi^{-1}) + \vec{u_2}^T \vec{u_2} \Sigma_2^{-1} - \vec{u_\phi} \vec{u_\phi} \Sigma_\phi^{-1}$$

if $\Sigma_2 = \Sigma_\phi = \Sigma$

$$= -2 x_i \Sigma^{-1} (u_2 - u_\phi) + \Sigma (\vec{u_2}^T \vec{u_2} - \vec{u_\phi}^T \vec{u_\phi}).$$

on substituting.

$$\log\left(\frac{\alpha}{1-\alpha}\right) + \frac{1}{2} \log\left(\frac{\Sigma}{\Sigma}\right) - 2 x_i \Sigma^{-1} (u_2 - u_\phi) + \Sigma (\vec{u_2}^T \vec{u_2} - \vec{u_\phi}^T \vec{u_\phi}).$$

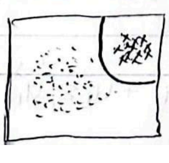
Is linear of the form $Ax + B = 0$

where

$$A = -2 \Sigma^{-1} (\vec{u_2} - \vec{u_\phi}).$$

$$B = \log\left(\frac{\alpha}{1-\alpha}\right) + \Sigma (\vec{u_2}^T \vec{u_2} - \vec{u_\phi}^T \vec{u_\phi}).$$

Let . & x be 2 classes with x being more concentrated around its mean



$\Rightarrow$ quadratic decision boundary.