Week 8 Exercise: Unsupervised Learning

Note: An indicative mark is in front of each question. The total mark is 10. You may mark your own work when we release the solutions.

$\boxed{1}$ 1. We have a 24-bit colour image of size 100×100. How many possible images of this size and bit depth?

Solution:
Possible number of images = (Possible number of pixel values)$^{100\times 100}$
Possible number of pixel values = 2^{24}
Possible number of images = $2^{24\times 100\times 100}$

$\boxed{3}$ 2. An alternative to derive PCA is to minimize the reconstruction error (Slide 26) for all N data samples $\mathbf{x}^{(i)}, i = 1, \cdots, N$, assuming that the mean $\boldsymbol{\mu} = \sum_i \mathbf{x}^{(i)}$ is zero. Take this approach to derive the first principal component (as the first eigenvector of the data matrix).

Solution: The most elegant proof is from `https://people.eecs.berkeley.edu/~jordan/courses/294-fall09/lectures/dimensionality/paper-1x2.pdf`.

Let us denote an **orthonormal** projection vector as $\mathbf{u}$. It will project an input vector $\mathbf{x}$ to a scalar $y = \mathbf{u}^\top\mathbf{x}$. Using this scalar to reconstruct $\mathbf{x}$ as $\hat{\mathbf{x}} = \mathbf{u}y = \mathbf{u}\mathbf{u}^\top\mathbf{x}$.

$$
\begin{aligned}
& \text{Reconstruction error} \\
= & \sum_{i=1}^{N} \left\| \mathbf{x}^{(i)} - \hat{\mathbf{x}}^{(i)}) \right\|^2 && (1)\\
= & \sum_{i=1}^{N} \left\| \mathbf{x}^{(i)} - \mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} \right\|^2 && (2)\\
= & \sum_{i=1}^{N} \left(\mathbf{x}^{(i)} - \mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)})^\top(\mathbf{x}^{(i)} - \mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} \right) && (3)\\
= & \sum_{i=1}^{N} \left(\mathbf{x}^{(i)\top} - \mathbf{x}^{(i)^\top}\mathbf{u}\mathbf{u}^\top)(\mathbf{x}^{(i)} - \mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} \right) && (4)\\
= & \sum_{i=1}^{N} \left(\mathbf{x}^{(i)\top}\mathbf{x}^{(i)} - \mathbf{x}^{(i)^\top}\mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} - \mathbf{x}^{(i)\top}\mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} + \mathbf{x}^{(i)^\top}\mathbf{u}\mathbf{u}^\top\mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} \right) && \text{note } \mathbf{u}^\top\mathbf{u}=1\\
= & \sum_{i=1}^{N} \left(\mathbf{x}^{(i)\top}\mathbf{x}^{(i)} - \mathbf{x}^{(i)^\top}\mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} \right) && (5)\\
= & \text{ constant} - \sum_{i=1}^{N} \left(\mathbf{x}^{(i)^\top}\mathbf{u}\mathbf{u}^\top\mathbf{x}^{(i)} \right) && (6)\\
= & \text{ constant} - \sum_{i=1}^{N} \left(\mathbf{u}^\top\mathbf{x}^{(i)} \right)^2 && (7)
\end{aligned}
$$

Note $\mathbf{u}^\top\mathbf{x}^{(i)}$ is the projection $y^{(i)} = \mathbf{u}^\top\mathbf{x}^{(i)}$ so the summation in Eqn. (7) is the variance. Maximising the variance minimises the reconstruction error so we have the same solution as that by variance maximisation.

$\boxed{1}$ 3. In k-means clustering, how could we determine k if it is not given?

Solution:
We can use the 'elbow plot' method. In which we plot k against the compactness measure it produces, and look for the 'elbow' of the plot. Typically, there will be a point where the graph will start plateauing in terms of the measure.
We calculate the measure by summing the squared distance between each point and the centroid of the cluster it belongs to.

$\boxed{2}$ 4. For the graph on Slide 39, compute the normalised cut $Ncut(A, B)$.

Solution:
$Ncut(A, B) = cut(A, B)\frac{Vol(A)+Vol(B)}{Vol(A)Vol(B)}$
$cut(A, B) = 0.1 + 0.2 = 0.3$
$Vol(A) = (0.8 + 0.6 + 0.1) + (0.8 + 0.8) + (0.8 + 0.6 + 0.2) = 4.7$
$Vol(B) = (0.8 + 0.2 + 0.7) + (0.8 + 0.7) + (0.8 + 0.8 + 0.1) = 4.9$
$Ncut(A, B) = 0.3 * (4.7 + 4.9)/(4.7 * 4.9) \approx 0.125$

$\boxed{3}$ 5. In spectral clustering, show that the smallest eigenvalue for the formulated generalized eigenvalue problem on Slide 41 is 0 with the corresponding generalized eigenvector $\mathbf{y} = \mathbf{1}$, hence the same "representation/embedding" for all nodes.

Solution:
$(D - W)y = \lambda Dy$
$(D - W)y = \lambda D^{\frac{1}{2}}D^{\frac{1}{2}}y$
$D^{\frac{-1}{2}}(D - W)y = \lambda D^{\frac{-1}{2}}D^{\frac{1}{2}}D^{\frac{1}{2}}y$
$D^{\frac{-1}{2}}(D - W)D^{\frac{-1}{2}}D^{\frac{1}{2}}y = \lambda ID^{\frac{1}{2}}y$
Make the substitution of $z = D^{\frac{1}{2}}y$
$D^{\frac{-1}{2}}(D - W)D^{\frac{-1}{2}}z = \lambda z$
If we set y to $\mathbf{1}$ we get
$z = D^{\frac{1}{2}}\mathbf{1}$
$D^{\frac{-1}{2}}(D - W)D^{\frac{-1}{2}}D^{\frac{1}{2}}\mathbf{1} = \lambda D^{\frac{1}{2}}\mathbf{1}$
$D^{\frac{-1}{2}}(D - W)I\mathbf{1} = \lambda D^{\frac{1}{2}}\mathbf{1}$
If we observe $(D - W)\mathbf{1}$, we can see that it's a summation of the rows of $D - W$
In a row of the Laplacian, $D - W$, we have the degree of the ith node, d_i on the diagonal, and all of the negative weights of the edges connected to node i filling the rest of the row. Therefore, adding across a row gives us:
$d_i + \sum_{j=1}^{n}(-w_{i,j}) = \sum_{j=1}^{n} w_{i,j} + \sum_{j=1}^{n}(-w_{i,j}) = 0$
Which means $(D - W)\mathbf{1} = \mathbf{0}$ and therefore the eigenvector corresponds to the eigenvalue $\lambda = 0$.