<u>CSCI 347 Cheat Sheet: Dimensionality Reduction</u>

Volume of a sphere in d dimensions with radius r: is given by: $\left(\frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}\right) r^d$

where:

$$\Gamma\left(\frac{d}{2}+1\right) = \begin{cases} \left(\frac{d}{2}\right)! & \text{if } d \text{ is even} \\ \sqrt{\pi} & \text{if } d \text{ is odd} \end{cases}$$

As dimensionality increases, the volume of a sphere with fixed radius increases for a while, then decreases. This is an example of how our intuition for low-dimensional spaces fails when we try to apply it to higher dimensions.

Feature selection is a form of dimensionality reduction, where a set of k attributes (features) is selected from the original set of $d > k$ attributes of a data set. A simple but useful way to select the k attributes is by finding those with greatest variance.

Matrix-vector multiplication:
Given a matrix $A \in R^{n \times d}$ and a vector $x \in R^d$, multiplying A by x on the left results in the vector Ax, with entries corresponding to the dot product of each row of A with x :

$$Ax = \begin{pmatrix} A(1,:)x \\ A(2,:)x \\ A(3,:)x \\ \vdots \\ A(n,:)x \end{pmatrix}$$

We can think of the matrix as a **linear transformation** that can be applied to any *d*-dimensional vector x. Many matrices can be decomposed into matrix factors that correspond to either a **rotation** or a **scaling** of a vector.

Eigenvectors and eigenvalues:
Given a matrix $A \in R^{n \times n}$, we can find **eigenvectors** with corresponding **eigenvalues** such that $: Av = \lambda v$
Thus we can think of this set of vectors as those for which the **linear transformation** of v corresponding to multiplication by A is just a scaling of v by a constant factor λ.

Principal component analysis (PCA):
An algorithm that linearly transforms a data matrix *D* in a way that will eliminate covariance among attributes. More specifically, PCA creates new attributes using **linear combinations** of the original attributes. We can think of these linear combinations of the original attributes as **new coordinate axes**. The linear transformation that PCA applies (the matrix that we multiply the data by) is the matrix of **eigenvectors** of the **covariance matrix** of the data. This transformation is optimal in the sense that after the linear transformation**, the variance along the new coordinate axes will be maximal** and there is no covariance among the new coordinate axes.

Principal component analysis (PCA) algorithm:

PCA(D = {x₁, …, xₙ}, α):

1. $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $Z = D - 1 \cdot \mu^T$
3. $\Sigma = \frac{1}{n-1} Z^T Z$
4. $(\lambda_1, \lambda_2, \ldots, \lambda_d) = \text{eigenvalues}(\Sigma)$
5. $(u_1, u_2, \ldots, u_d) = \text{eigenvectors}(\Sigma)$
6. Choose the smallest r such that $f(r) \geq \alpha$, where:
$$f(r) = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$
7. Find the new coordinates a_i:
$U_r = (u_1 \quad \cdots \quad u_r)$, $\quad a_i = U_r^T x_i$
for $i = 1, \ldots, n$