CS 189: Introduction to Machine Learning - Discussion 1

1. Probability Review

   There are n archers all shooting at the same target (bullseye) of radius 1. Let the score for a particular archer be defined to be the distance away from the center (the lower the score the better, and 0 is the optimal score). Each archer's score is independent of the others, and is distributed uniformly between 0 and 1. What is the expected value of the worst score?

   Solution: For this problem, we compute the CDF, take the derivative to get the PDF, then calculate the expectation. We defined a random variable $Z = \max\{X_1, \ldots, X_n\}$.

   - Computing the CDF:
     $F(z) = P(Z < z) = P(X_1 < z)P(X_2 < z) \ldots P(X_n < z) = \prod_{i=1}^{n} P(X_i < z)$
     $$= \begin{cases} 0 & \text{if } z < 0 \\ z^n & \text{if } 0 \leq z \leq 1 \\ 1 & \text{if } z > 1 \end{cases}$$
   - Getting the PDF:
     $$P(Z = z) = \frac{d}{dz}F(z) = \begin{cases} nz^{n-1} & \text{if } 0 \leq z \leq 1 \\ 0 & \text{otherwise} \end{cases}$$
   - Computing $E[Z]$:
     $$E[Z] = \int_0^1 znz^{n-1}\mathrm{d}z = n\int_0^1 z^n \mathrm{d}z = \frac{n}{n+1}\left[z^{n+1}\right|_0^1 = \frac{n}{n+1}$$

2. Maximum Likelihood Estimation

   Given N i.i.d. Poisson random variables, $x_1, x_2, ..., x_N$, find the maximum likelihood estimator for the parameter of the distribution, λ. Recall for a Poisson R.V., $p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$.

Solution:

$$\begin{aligned}
L(x_1, x_2, ..., x_N; \lambda) &= p(x_1, x_2, ..., x_N; \lambda) \\
&= p(x_1; \lambda) p(x_2; \lambda) \cdots p(x_N; \lambda) \\
&= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda}\lambda^{x_N}}{x_N!} \\
&= \frac{e^{-N\lambda}\lambda^{\sum x_i}}{x_1! x_2! \cdots x_N!}
\end{aligned}$$

Taking the log-likelihood:

$$\ln L(x_1, x_2, ..., x_N; \lambda) = -N\lambda + (\ln \lambda) \sum x_i - \ln \prod x_i!$$

Taking the derivative and setting to zero (noting that the expression above is concave in λ):

$$\frac{d(\ln \text{Ł})}{d\lambda} = -N + \frac{\sum x_i}{\lambda} = 0$$

and hence

$$\hat{\lambda} = \frac{\sum x_i}{N}$$

3. Linear Algebra

Find the eigenvalues and corresponding eigenvectors of the following matrix.

$$A = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}$$

Solution: Remember that an eigenvector is a vector $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$, where the constant λ is the eigenvalue corresponding to $\vec{v}$. We manipulate the above equation to be $(A - \lambda I)\vec{v} = 0$, which implies that $A - \lambda I$ is a singular matrix since it has an eigenvalue of 0.

$$A - \lambda I = \begin{bmatrix} 2-\lambda & 1 \\ 0 & -1-\lambda \end{bmatrix}$$

We can take the determinant of the above matrix and set it to zero in order for the matrix to be singular, giving us the following characteristic polynomial:

$$-(2-\lambda)(1+\lambda) = 0$$
$$\lambda_1 = 2, \lambda_2 = -1$$

We can place each λ back into $A - \lambda I$ and solve for the corresponding eigenvectors. Note that the eigenvectors can be scaled up or down arbitrarily by a constant factor.

$$(A - \lambda_1 I)\vec{v}_1 = \begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix} \vec{v}_1 = 0 \rightarrow \vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Likewise for $\vec{v}_2$

$$(A - \lambda_2 I)\vec{v}_2 = \begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix} \vec{v}_2 = 0 \rightarrow \vec{v}_2 = \begin{bmatrix} -\frac{1}{3} \\ 1 \end{bmatrix}$$

4. Projections

Given a plane $x + y + z = 4$ and Point A located at $(2, 6, 8)$, find coordinates of the closest Point B on the plane to Point A. How far away is Point A from Point B?

Solution: Consider the normal vector $[1, 1, 1]$ of the plane defined by $x+y+z = 4$, which is by definition perpendicular to the surface. Note that the vector between Points A and B ($\overrightarrow{BA}$) will also be perpendicular to the plane. Thus, we can follow the normal vector from Point A until we reach the plane, and the point of intersection will be the location of Point B. Therefore: we define w as a constant attached to the normal vector. Note that if the normal vector had L2 Norm 1, then w would be the distance from point A to point B.

$$([2, 6, 8] - w[1, 1, 1]) \cdot [1, 1, 1] = 4$$
$$[2 - w, 6 - w, 8 - w] \cdot [1, 1, 1] = 4$$
$$16 - 3w = 4$$

$$w = 4$$

Coordinates of Point B:

$$[2, 6, 8] - 4[1, 1, 1] = [-2, 2, 4]$$

Lets sanity check our answer by choosing another arbitrary point C (say $(0, 0, 4)$) on our plane and checking to see if $\overrightarrow{BA} \perp \overrightarrow{BC}$:

$$([2, 6, 8] - [-2, 2, 4]) \cdot ([0, 0, 4] - [-2, 2, 4])$$
$$= [4, 4, 4] \cdot [2, -2, 0] = 8 - 8 = 0$$

Since now we know the coordinates of Point B, we can use the Euclidean Distance formula to determine the L2 Norm of $\overrightarrow{BA}$:

$$\sqrt{(2+2)^2 + (6-2)^2 + (8-4)^2} = 4\sqrt{3}$$