CMPE 302 - Python Exercise Worksheet

Anıl Demirel

March 24, 2025

Question 1 - Ridge Regression and Bias-Variance Tradeoff

Given the dataset:

$$X = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} 2.1 \\ 2.9 \\ 3.9 \\ 5.2 \end{bmatrix}$$

- Compute the Ridge Regression weights using the formula:

$$W = (X^T X + \delta^2 I)^{-1} X^T Y$$

- Perform this calculation in Python for $\delta = 10$
- Plot how weights change when δ increases.

Question 2 - Maximum Likelihood Estimation (MLE) with Gaussian Noise

Generate the dataset using Python:

```
np.random.seed(0)
X = np.linspace(0, 10, 50)
Y = 2.5 * X + np.random.normal(0, 2, 50)
```

- Compute the MLE estimate for W by minimizing:

$$\sum_{i=1}^{N} (y_i - wx_i)^2$$

- Plot your linear fit and compare it to the true slope 2.5.

Question 3 - Entropy, Cross Entropy and KL-Divergence

Given the probability distributions:

$$P = [0.5, 0.5] \quad \text{(Fair coin)}, \quad Q = [0.3, 0.7] \quad \text{(Model prediction)}$$

- Write Python code to compute the entropy $H(P)$:

$$H(P) = -\sum p_i \log_2 p_i$$

- Compute the cross-entropy $CE(P, Q)$:

$$CE(P, Q) = -\sum p_i \log_2 q_i$$

- Compute the KL-Divergence:

$$KL(P||Q) = CE(P, Q) - H(P)$$

Question 4 - Maximum A Posteriori Estimate (MAP)

Generate synthetic data:

```
np.random.seed(0)
X = np.linspace(0, 5, 30)
Y = 3 * X + np.random.normal(0, 1, 30)
```

- Compute the MAP estimate for W by minimizing:

$$W_{MAP} = \arg\min_W \left(\sum_{i=1}^{N} (y_i - wx_i)^2 + \lambda W^2 \right)$$

- Compute results for $\lambda = 1$.
- Compare the result of MAP with the MLE estimate.

Bonus - Bias-Variance Tradeoff Visualization

- Generate a random dataset
- Fit polynomial models of degree 1, 3, and 10
- Plot and observe the training and test errors