

HOMWORK 1

MLE, MAP ESTIMATES; LINEAR AND LOGISTIC REGRESSION

CMU 10-701: MACHINE LEARNING (SPRING 2017)

OUT: Jan 31

DUE: Feb 10, 11:59 PM

NAME: Mengwen He

ADREW ID: mengwenh

Part A: Multiple Choice Questions

- For each case listed below, what type of machine learning problem does it belong to?
 - Advertisement selection system, which can predict the probability whether a customer will click on an ad or not based on the search history.
Answer:
 - U.S post offices use a system to automatically recognize handwriting on the envelope.
Answer:
 - Reduce dimensionality using principal components analysis (PCA).
Answer:
 - Trading companies try to predict future stock market based on current market conditions.
Answer:
 - Repair a digital image that has been partially damaged.
Answer:

Type of machine learning problem:

- Supervised learning: Classification
 - Supervised learning: Regression
 - Unsupervised learning
- For four statements below, which one is wrong?
 - In maximum a posterior (MAP) estimate, data overwhelms the prior if we have enough data.
 - There are no parameters in non-parametric models.
 - $P(X \cap Y \cap Z) = P(Z|X \cap Y)P(Y|X)P(X)$.
 - Compared with parametric models, non-parameter models are flexible, since they don't make strong assumptions.

Answer:

- There are about 12% people in U.S. having breast cancer during their lifetime. One patient has a positive result for the medical test. Suppose the sensitivity of this test is 90%, meaning the test will be positive with probability 0.9 if one really has cancer. The false positive is likely to be 2%. Then what is the probability this patient actually having cancer based on Bayes Theorem?
 - 90%
 - 61%
 - 38%
 - 11%

Answer:

- What is the most suitable error function for gradient descent using logistic regression?

- A. The negative log-likelihood function
- B. The number of mistakes
- C. The squared error
- D. The log-likelihood function

Answer:

Part B, Problem 1: Bias-Variance Decomposition

Consider a p -dimensional vector $\vec{x} \in \mathbb{R}^p$ drawn from a Gaussian distribution with an identity covariance matrix $\Sigma = I_p$ and an unknown mean $\vec{\mu}$, i.e. $\vec{x} \sim \mathcal{N}(\vec{\mu}, I_p)$. Our goal is to evaluate the effectiveness of an estimator $\hat{\vec{\mu}} = f(\vec{x})$ of the mean from only a single sample (i.e. $n = 1$) by measuring its mean squared error $\mathbb{E}[\|\hat{\vec{\mu}} - \vec{\mu}\|^2]$, where $\|\cdot\|^2$ is the squared Euclidean norm and the expectation is taken over the data generating distribution.

Part C, Problem 2: Linear Regression

Part D, Problem 3: MLE, MAP and Logistic Regression

Part E, Problem 3: Programming Exercise