MECE 6616 - Robot Learning

Midterm Review

Spring 2020

1. Clustering
    a. Define within-cluster scatter and between-cluster scatter. Show that minimizing within-cluster scatter is equivalent to maximizing between-cluster scatter.
    b. Write the K-means clustering algorithm in pseudocode.
    c. Explain the following metrics for determining how similar two clusters are: a) closest neighbor; b) furthest neighbor; c) average.
    d. What is Vector Quantization, and why is it usable as a compression algorithm?
2. Dimensionality Reduction
    a. Assume a dimensionality reduction mapping $\mathbf{x}_i \rightarrow f(\mathbf{y}_i)$ where $\mathbf{x}_i \subset \Re^d$, $\mathbf{y}_i \subset \Re^m$, $m << d$. What is a good quantitative measure of whether $f()$ constitutes a "good" mapping?
    b. Assume a linear mapping $\mathbf{x}_i = \mathbf{A}\mathbf{y}_i$, where $\mathbf{x}_i \subset \Re^d$, $\mathbf{y}_i \subset \Re^m$, $\mathbf{A} \subset \Re^{d \times m}$. Where can we find the basis vectors of the lower dimensional subspace that is being used?
    c. Assume a set of data points $\mathbf{x}_i \subset \Re^d$, $i=1..N$. How do I determine the optimal linear mapping into a 2-dimensional subspace that minimizes my reprojection error into the original d-dimensional space? You can use an approach based either on Eigenvalue decomposition or Singular Value Decomposition.
3. The Kernel Trick: Consider a non-linear projection of a data point into a higher dimensional space $\mathbf{x}_i \rightarrow \Phi(\mathbf{x}_i)$.
    a. What is the Gram matrix $\mathbf{K}$ of high dimensional projections $\Phi(\mathbf{x}_i)$, and what are its dimensions?
    b. What is the definition of a kernel function, and what property must it have in order to be useful?
    c. What is the expression of the Polynomial kernel function?
4. Linear Regression: Assume a linear mapping between features $\mathbf{x} \subset \Re^{d1}$ and labels $y \subset \Re$ of the form $\mathbf{x}_i^T \mathbf{w} = y_i$.
    a. What is the definition of the optimal $\mathbf{w}$ if we are trying to achieve a least squares fit?
    b. How is the optimal least squares $\mathbf{w}$ computed?
    c. What is the objective being optimized through Ridge Regression?

    d. Describe the regularization parameter used in conjunction with Ridge Regression, and explain its role in the objective being optimized.
5. Classification
    a. Assume a 2-class classification problem. How do I compute a decision boundary via Linear Regression?
    b. What does a maximum margin classifier do when applied to a two-class problem where the two classes are linearly separable?
    c. What is the main difference between a Support Vector Machine and a linear classifier learned via linear regression, in terms of the classification boundary each one produces?
    d. What is a Support Vector, in the context of Support Vector Machines?
    e. Describe at least one regularization parameter used in conjunction with Support Vector Machines, and explain its role in the objective being optimized.
    f. Outline in pseudocode a greedy algorithm used to partition the feature space for a Decision Tree.
    g. What is the most common model that a regression Decision Tree fits to the data in each of its leaves?
    h. How does a Decision Tree make a regression prediction? What about a classification prediction?
    i. Explain how a Random Forest is different from a Decision Tree, and what are its advantages.
6. Computer Vision fundamentals
    a. Name three properties of Perspective Projection.
    b. What is the definition of the camera matrix, and what are its dimensions?
    c. Name three of the factors that determine the amount of light received by a given pixel of a camera sensor.
    d. What is the “correspondence problem” that must be solved in order to determine scene geometry via stereo vision?
    e. Describe three methods that can be used to determine the “depth” of a point imaged by a camera (a.k.a. the location of the point in three-dimensional space).
    f. Explain why PCA is an appropriate method for finding the plane that best fits all the 3D points in a point cloud. Also explain why this algorithm is also not well suited for finding a planar surface in a point cloud of a cluttered scene.
    g. Outline in pseudocode the RANSAC algorithm. Explain why it is well suited for finding a planar surface in a point cloud of a cluttered scene.
7. Learning on Vision example: Bag of Features
    a. What is the main limitation for applying classification via linear regression or Support Vector Machines directly on raw image data?
    b. What is a keypoint detector?
    c. What is a keypoint descriptor?
    d. What are desirable features of a keypoint descriptor?
    e. How is the Codeword dictionary created?

f. Assuming a Codeword dictionary has been created, how is an image encoded into a feature vector?