Week 7 Lecture Notes: Support Vector Machines

**Optimization Objective:**

* **Support Vector Machine (SVM)**: A type of supervised machine learning algorithm. Advantageous for being clean and powerful.
  + Logistic Regression:
    - For y=1y=1: hθ(x)≈1hθ​(x)≈1 and ΘTx>>0ΘTx>>0
    - For y=0y=0: hθ(x)≈0hθ​(x)≈0 and ΘTx<<0ΘTx<<0
  + **Modifications in Cost Function for SVM**:
    - Instead of sigmoid curve in logistic regression, SVM uses a straight line for the costs.
    - cost1(z)=max⁡(0,k(1−z))cost1​(z)=max(0,k(1−z)) for y=1y=1
    - cost0(z)=max⁡(0,k(1+z))cost0​(z)=max(0,k(1+z)) for y=0y=0
    - **SVM Cost Function**: J(θ)=C∑i=1m[y(i)cost1(θTx(i))+(1−y(i))cost0(θTx(i))]+12∑j=1nΘj2J(θ)=C∑i=1m​[y(i)cost1​(θTx(i))+(1−y(i))cost0​(θTx(i))]+21​∑j=1n​Θj2​
  + **Interpretation of Hypothesis**:
    - In SVM: hθ(x)=1hθ​(x)=1 if ΘTx≥0ΘTx≥0 and hθ(x)=0hθ​(x)=0 otherwise. Not viewed as a probability.

**Large Margin Intuition:**

* SVMs are often thought of as **Large Margin Classifiers**.
  + For y=1y=1: We want ΘTx≥1ΘTx≥1
  + For y=0y=0: We want ΘTx≤−1ΘTx≤−1
  + When CC is large, SVMs separate positive and negative examples with a large margin.
  + **Large Margin**: Decision boundary is as far from both positive and negative examples as possible. Achieved when CC is large.
  + **Impact of CC**:
    - Large CC: Lower bias, high variance (smaller margin).
    - Small CC: High bias, low variance (large margin).

**Mathematics Behind Large Margin Classification:**

* **Vector Inner Product**:
  + Length of a vector: ∣∣v∣∣∣∣v∣∣
  + Projection of vv onto uu: p=uTvp=uTv
  + uTv=p×∣∣u∣∣uTv=p×∣∣u∣∣
  + In SVMs: ΘTx(i)=p(i)×∣∣Θ∣∣ΘTx(i)=p(i)×∣∣Θ∣∣
  + Large margin is due to SVM requiring projections p(i)p(i) to be large.

**Kernels:**

* **Kernels** allow SVMs to form non-linear decision boundaries.
  + Given x, compute new features based on proximity to landmarks l(i)l(i).
  + **Gaussian Kernel**: fi=similarity(x,l(i))=exp⁡(−∣∣x−l(i)∣∣22σ2)fi​=similarity(x,l(i))=exp(−2σ2∣∣x−l(i)∣∣2​)
  + The Gaussian kernel measures the "similarity" between a given example xx and landmarks l(i)l(i).

**Key Takeaways:**

1. **SVMs** provide a different way to think about classification compared to logistic regression. They aim to find the largest possible margin to separate data.
2. SVMs can be modified to work in non-linear scenarios through the use of **kernels**, which compute similarities with respect to landmarks in the dataset.
3. The regularization parameter **CC** in SVM plays a similar role as 1λλ1​ in logistic regression. It controls the balance between achieving a large margin and ensuring that the SVM fits the training data well.
4. **Mathematically**, SVMs can be understood using vector operations, where the decision boundary is represented in terms of a vector ΘΘ that is perpendicular to the boundary itself.
5. Overall, SVMs provide a robust way to perform classification, especially in scenarios where clear margins can be identified between different classes in the data.