MATH 156: Machine Learning — Project #3

This is due on November 4 by 11:59 pm (PST).

1. Consider the model $y(\boldsymbol{x}, \boldsymbol{w}) = \text{sign}(\boldsymbol{w}^\top \boldsymbol{x})$. Let the single data point $\boldsymbol{x}_1$ with target value $t_1 = 1$ be given. Show that the set of all weight vectors $\boldsymbol{w}$ that define a model which correctly classifies this point is convex. Hint: First, write down what it means that $\boldsymbol{w}$ correctly classifies $\boldsymbol{x}_1$. Assume this is true for $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. Now show it holds for $\lambda \boldsymbol{w}_1 + (1-\lambda)\boldsymbol{w}_2$ where $\lambda \in [0, 1]$.

   Solution: For w to be correctly classified by x_1 means $w_1^T x_1 \geq 0$, as this implies $y(x_1, w_1) = 1$ and the target value $t_1 = 1$ is given. Assume this is true for both w_1 and w_2 s.t. w_1, $w_2 \in w$ where w is the weight space of the set of all weight vectors that correctly classifies the point x_1.
   Therefor any point $\vec{w}$ on the line between w_1 and w_2 is given by:

$$\vec{w} = \lambda \boldsymbol{w}_1 + (1-\lambda)\boldsymbol{w}_2$$

   where

$$\lambda \in [0, 1]$$

$$\implies y_k(\vec{w}) = y_k(\lambda \boldsymbol{w}_1 + (1-\lambda)\boldsymbol{w}_2) = \lambda y_k(\boldsymbol{w}_1) + (1-\lambda) y_k(\boldsymbol{w}_2)$$

$$\text{Since } \boldsymbol{w}_1,\ \boldsymbol{w}_2 \in \mathbb{R}^k\ y_k(\boldsymbol{w}_1) > y_j(\boldsymbol{w}_2) \text{ for all j} \neq \text{k}$$

   and

$$y_k(\boldsymbol{w}_2) > y_j(\boldsymbol{w}_1) \text{ for all j} \neq \text{k}$$

$$\implies y_k(\boldsymbol{w}) = \lambda y_k(\boldsymbol{w}_1) + (1-\lambda) y_k(\boldsymbol{w}_2)$$

   and

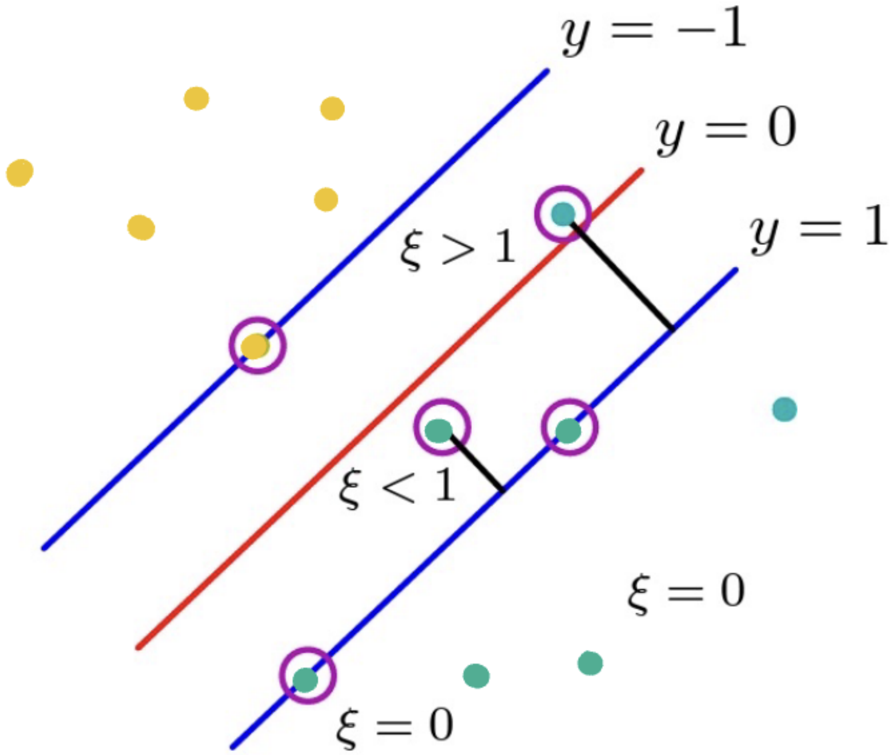
$$y_k(\boldsymbol{w}) > \lambda y_j(\boldsymbol{w}_1) + (1-\lambda) y_j(\boldsymbol{w}_2) = y_j(\boldsymbol{w}), \forall j \neq k$$

   Therefor the set of all weight vectors w is convex.

2. Consider the soft-margin SVM problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{n=1}^{N}\xi_n \\ &\text{s.t. } t_n(\boldsymbol{w}^T\phi(\boldsymbol{x}_n)+b) \geq 1-\xi_n \quad \forall n \\ &\xi_n \geq 0 \quad \forall n \end{aligned}$$

Sketch a two-dimensional two-class toy example and answer the following geometrically:



(a) Where does a data point lie relative to where the margin is when $\xi_n = 0$? Is this data point classified correctly?
Solution: The point will lie on the support vector. This point is classified correctly.

(b) Where does a data point lie relative to where the margin is when $0 < \xi_n \leq 1$? Is this data point classified correctly?
Solution: This point will lie on the other side of the margin boundary, but lie in the margin area between the support vector and the decision boundary. This means the point is still classified correctly, even though it is on the "wrong" side of the margin boundary.

(c) Where does a data point lie relative to where the margin is when $\xi_n > 1$? Is this data point classified correctly?
Solution: This point lies on the wrong side of both the margin boundary as it exceeds the margin magnitude and also goes beyond the decision boundary. This point will lie in the margin of the other class and be misclassified. Therefor this point is not correctly classified.

3. The derivative of the logistic sigmoid activation function can be expressed in terms of the function value itself,

$$\frac{\partial \sigma(a)}{\partial a} = \sigma(a)(1 - \sigma(a)).$$

Derive the corresponding result for the hyperbolic tangent function, $\tanh(a)$,

$$\frac{\partial \tanh(a)}{\partial a} = 1 - \tanh^2(a).$$

Solution:

$$\frac{\partial \tanh(a)}{\partial a} = \frac{\partial}{\partial a} \frac{\text{sinh(a)}}{\text{cosh(a)}}$$

$$= \frac{\frac{\partial}{\partial a}\text{sinh(a)} \cdot \text{cosh(a)} - \frac{\partial}{\partial a}\text{cosh(a)} \cdot \text{sinh(a)}}{\cosh^2(a)}$$

$$= \frac{\cosh^2(a) - \sinh^2(a)}{\cosh^2(a)}$$

$$= 1 - \frac{\sinh^2(a)}{\cosh^2(a)}$$

$$= 1 - \tanh^2(a)$$

4. Ethical implications of data science:
    - Data Collection and Security:
        - Training data should be acquired with consent
        - Personal/sensitive data should be stored/transmitted securely
    - Composition of data sets:
        - Data set should match diversity of target population
        - Data sets scraped from internet need to be checked for explicit bias
    - Non-discriminatory decision-making:
        - Cannot base decision upon protected attributes
        - Proxies or implicit biasin training data can lead to unequal outcomes

    Skim through the publicly-available paper[1] and summarize it, or watch the documentary "Coded Bias" and summarize it, or skim through the publicly-available survey paper[2] and pick a type of bias and describe it.

    Summary of Coded Bias: A MIT Researcher that was using Computer Vision Technology at the MIT Media Lab for a school project using computer vision. Her project involved creating an augmented reality mirror that could project people or images onto your face. However, the researcher had trouble getting her face to work with the system. It turns out that when she would put on a white mask, the computer vision software would work. The researcher realized that there was a bias in the training set of data. The dataset was skewed towards men and white individuals. Largely skewed datasets leads to skewed results. It turns out that inequalities within society are imprinted within the data leading to discrimination by AI algorithms. The data embeds the past. This becomes a problem when machine learning algorithms are being touted as true unbiased solution using the math as a shield for corrupt practices. There is a need to monitor for bias in big data.

    Currently we have narrow AI, not generalized AI. Early machine learning researchers believed that computer intelligence could be measured by ability to play games, specifically chess. However, this technology was being built by a small homogeneous group that embeds their own biases into the systems they build. The power is with those who own the code. Large corporations have begun to utilize AI to earn revenue with commercial applications. Amazon sorts through resumes for hiring, program was biased against women rejecting all their resumes. This is because their are very few women working in powerful tech jobs at amazon, the machine was only going off the data provided. machine learning can not be used to replicate the world today as there will not be social progress. Microsoft shut off their NLP AI Tay after just 16 hours online on twitter after it started becoming racist, sexist, and xenophobic. Amazons Rekognition CV openAPI was shown to have racial biases and has been sold and implemented by various US Law enforcement agencies.

    In the UK 98% of matches from their facial recognition systems are identifying innocent people as a wanted person. There exists laws about taking fingerprints, but none regarding bio-metric photos being stored in a database without consent. In China, protesters use lasers to confuse the cameras facial recognition software. 117 Million people in the United States have their faces in a database without checks for accuracy–start of a mass surveillance state without proper safeguards and checks in place.

    Not just Computer Vision, in social media feeds, ads that are displayed are all powered by AI, which shapes your worldview. Automated admissions or job/credit application processes, the world is becoming increasingly automated. People do not know how the algorithm reaches the decision. much of the actions people take on the internet now are recorded, logged, and analyzed and fed into learning algorithms to build a synopsis of who you are. All of this is done without proper safeguards in place. Internet advertising as data scientist predatory industries are competing for their attentions, this increases inequality. Companies can in essence predict who you are and what you like, even political affiliation. Facebook messages were used in order to swing the US presidential election in 2016, and the BREXIT vote. There is no oversight into the power that these companies hold.

    Machine Learning is hard to understand what is happening under the hood, and is thought of as a black box. A value added model fired a teacher who consistently one teacher of the year. The algorithm determined he

[1] Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

[2] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35.

was an ineffective teacher but was not told how. Machine learning models have also been given to judges when determining probation and parole terms, questions proxies for races and class. Black individuals got higher scores for reactivation than white individuals.

The researchers ultimately testified in front of congress about the effects that AI algorithms have and the need for effective safeguards to prevent bias within these systems. Algorithms that are disproportionately effective on white men, created and designed by one demographic. There is a need for oversight on the ethical collection, training composition, and decision making of AI systems.