SMAI-M20-03: Performamnce Metrics

C. V. Jawahar

IIIT Hyderabad

August 14, 2020

Recap

1. Recap/Repeat of L01
   - Scope, Course Plans.
   - L01: https://www.dropbox.com/s/ltmyx9y15hxnmm8/L1.pdf?dl=0
   - L02: https://www.dropbox.com/s/7216t0zl39xgmeq/L2.pdf?dl=0
2. Scope:
   - Representation as a vector in R^d
   - Learn a function $y = f(\mathbf{W}, \mathbf{x})$ from the data.
     - Notion of Training and Testing
   - Feature Transformation as a useful trick:

$$\mathbf{x}' = \mathbf{W}\mathbf{x}$$

     - Dimensionality Reduction
   - Two Simple Classification Schemes:
     - Nearest Neighbour Algorithm
     - Linear Classification
     - — $sign(\mathbf{w}^T\mathbf{x})$; Either +ve or -ve.
     - — Many ways to extend to more than 2 classes

This Lecture: Performance Metrics

- In the classification setting:
  - Accuracy, Confusion Matrix, Precision, Recall
- In the ranking/retrieval setting:
  - Precision, Recall, AP, F-score.
- Many more performance metrics see for example:
  - `https://en.wikipedia.org/wiki/Confusion_matrix`
- Pre-Class Videos in the Channel (12 min) (Please make sure you see it before the class)
- (*Let us spend 5 mins today to watch for those who have missed. But make sure from the next class you spend 20-30 mins before the lecture.*)

Blank

Discussions Point - I

Q: We have heard in recent months:

- The PCR-Test (or RT-PCR) for COVID (TEST-1) has some "false negatives". (PCR-Test tests whether a patient has COVID right now by looking at the swab.)
- The Antibody test used or COVID (TEST-II) has very high "false positives". (Antibody test is used to know the immunity level. It checks whether a person has already got COVID, (usually after recovered or infected unnoticed).)

What does these sentences mean? Can you please explain the implication of these two in two sentences S1 and S2 (say to your grand mother!). Support (like) the best explanations!!.
(Type S1 and S2 in chat section)

Blank

Discussions Point -II

Q: Let us consider that FN rate of TEST-I is 10% and the FP of rate of TEST-II is 50%.
An SMAI student gave a recommendation to the Govt that Every Person should be tested three times (say in a day) and majority label should be assigned.

- Is this student, technically sound?
- Why is that Govt is not ready to listen to his/her recommendation?

Blank

Discussion Point - III

Q: There was a proposal to an airport from a supplier:
"We have a multi-sensor product that can recognize terrorists reliably. We can classify a normal person vs terrorist (two class classification problem) with an accuracy of 99.5% "

- Is this a good product for detecting terrorists?
- As an SMAI student, who evaluate this product, what performance metrics you will prefer?

Blank

Discussions Points - IV

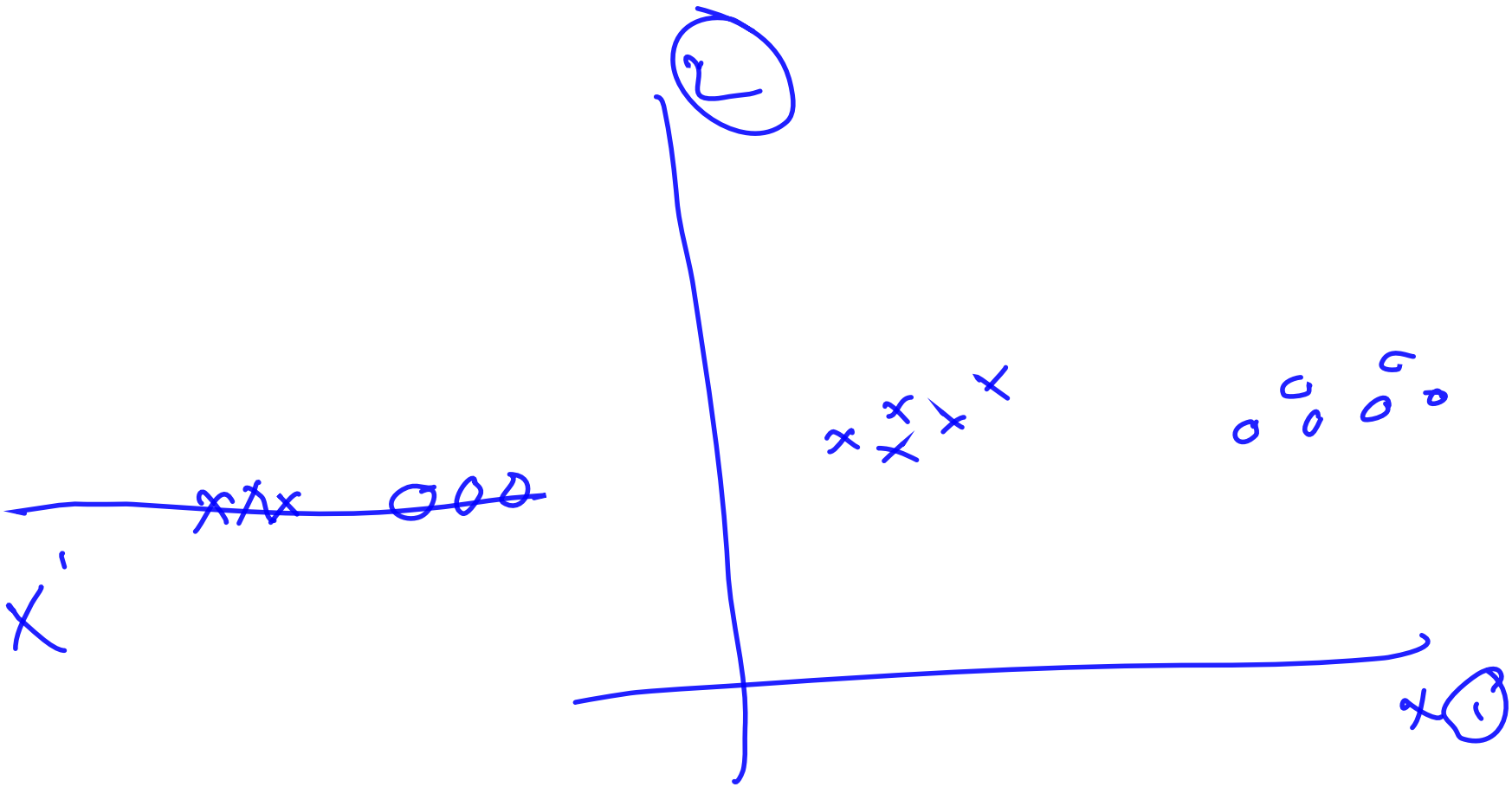
In a problem of classification of "cricket ball vs Tennis ball" there were two features (i) size/radius measured in micometer (ii) Weight measured in Tons.

- Fact: SMAI student felt that size should be in cm (1 cm $= 10^4$ mico meter) and weight in Kg (1 Ton $= 10^3$ Kg) for good performance of the algorithm she uses.
- Suggest the 2×2 matrix $\mathbf{W}$ to obtain the new feature representation from the old one.

$$\mathbf{x}' = \mathbf{W}\mathbf{x}$$

Blank

Blank



Review Question (one, none or more correct)

We are given a set of 2D points from two classes, as shown in the Figure.
Q: To make the computations efficient, we want to do a dimensionality reduction from 2D to 1 D with the help of a 1×2 matrix $\mathbf{W}$

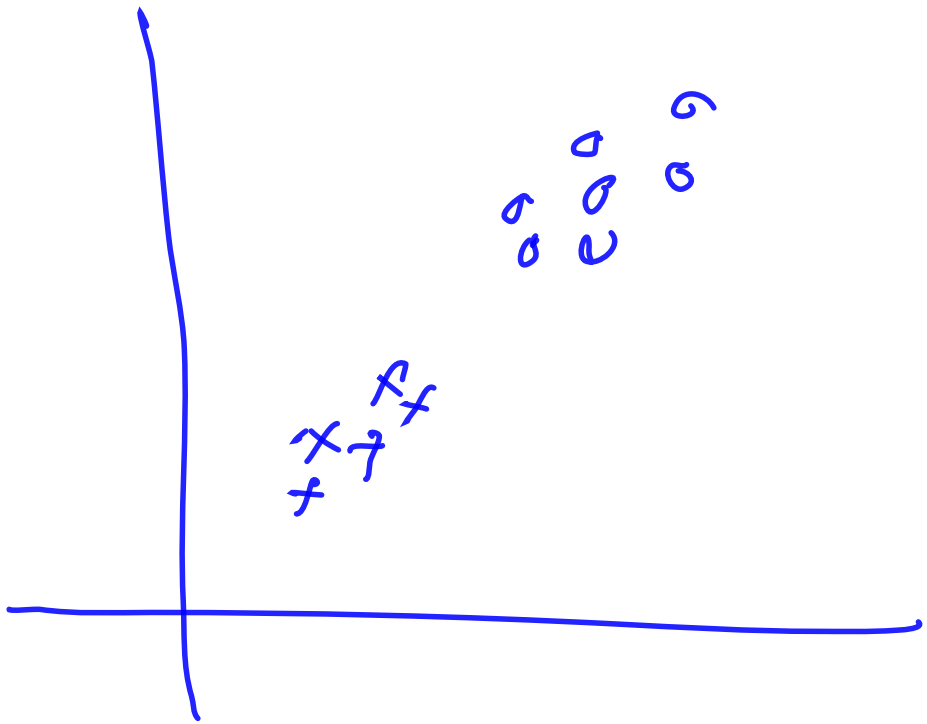
$$\mathbf{x}' = \mathbf{W}\mathbf{x}$$

What should be the $\mathbf{W}$ matrix be in this case?
(Indeed the goal is to get good classification performance in the new feature space $\mathbf{x}'$, while the computations could be efficient)

(a) $[1, 0]$ (b) $[-1, 0]$ (c) $[2, 0]$ (d) $[1, 1]$ (e) $[0, 1]$ (f) $[0, -1]$ (g) $[1, 0]^T$

Blank

Blank



Review Question (one, none or more correct)

We are given a set of 2D points from two classes, as shown in the Figure. We want to "rorate" the data so that points are spread across first (x) axis (i.e., something like rotate clockwise by 45°)

$$\mathbf{x}' = \mathbf{W}\mathbf{x}$$

What should be the 2×2 matrx $\mathbf{W}$ be in this case?

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ (g) Any one of the above (h) None of the above

Blank

Review Question (one, none or more correct)

What is the probability that, in a room of 30 people, there is a pair of people who have the same birthday. (use calculator, pick the closest)
(a) 0.11 (b) 0.31 (c) 0.51 (d) 0.71 (e) 0.91

What Next:?

- Logistics: Watch MS-Teams for use of “shiksha”.
  - Announcements Channel for General Announcements
  - Lecture related material on ”Lecture Announcements”
  - Emails at: smai.m2020@gmail.com
- Topics and Preparation:
  - Revise: Rank of a matrix. Interpretation of Rank.
  - Continue on Features and Supervised Learning (Next Week)
  - Budget time to watch videos before class.
- Next Week:
  - Start submission of regular HWs (a week to submit) and Class Reviews (in class 5-10 mins)