2. Feature Vectors Transformation

Note: The problems on this page appeared as ungraded earlier in Homework 1. They are graded here.

Consider a sequence of n-dimensional data points, $x^{(1)}, x^{(2)}, \ldots$, and a sequence of m-dimensional feature vectors, $z^{(1)}, z^{(2)}, \ldots$, extracted from the x's by a linear transformation, $z^{(i)} = Ax^{(i)}$. If m is much smaller than n, you might expect that it would be easier to learn in the lower dimensional feature space than in the original data space.

2. (a)

1.0/1 point (graded)

Suppose $n = 6, m = 2$, z_1 is the average of the elements of x, and z_2 is the average of the first three elements of x minus the average of fourth through sixth elements of x. Determine A.

Note: Enter A in a list format: $[[A_{11}, \ldots, A_{16}], [A_{21}, \ldots, A_{26}]]$

[[1/6,1/6,1/6,1/6,1/6,1/6] **Answer:** [[1/6,1/6,1/6,1/6,1/6,1/6], [1/3,1/3,1/3,-1/3,-1/3,-1/3]]

Solution:

- $A = \left[[1/6, 1/6, 1/6, 1/6, 1/6, 1/6], [1/3, 1/3, 1/3, -1/3, -1/3, -1/3]\right]$

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

2. (b)

1.0/1 point (graded)

Using the same relationship between z and x as defined above, suppose $h(z) = sign(\theta_z \cdot z)$ is a classifier for the feature vectors, and $h(x) = sign(\theta_x \cdot x)$ is a classifier for the original data vectors. Given a θ_z that produces good classifications of the feature vectors, determine a θ_x that will identically classify the associated x's.

Note: Use `trans(...)` for transpose operations, and assume A is a fixed matrix (enter this as `A`).

Note: Expects θ_x (an $[n \times 1]$ vector), not $\theta_x^\top$.

$\theta_x =$ trans(A)*theta_z **Answer:** trans(A)*theta_z

Solution:

From above, we have the relationship that $z = Ax$. Therefore $\theta_z \cdot z = \theta_z \cdot Ax = \theta_z^\top Ax = (A^\top \theta_z) \cdot x$. So take $\theta_x = A^\top \theta_z$ and we have the same classifier.

Submit You have used 1 of 5 attempts

Answers are displayed within the problem

2. (c)

1/1 point (graded)

Given the same classifiers as in (b), if there is a θ_x that produces good classifications of the data vectors, will there **always** be a θ_z that will identically classify the associated z's?

Note: A is a fixed matrix.

- Yes
- No ✔

Solution:

No. Here we provide a formal condition when there will be a θ_z that will identically classify the associated z's. Formally, suppose we are given a θ_x that correctly classifies the points in data space of dimension $m < n$. We are looking for θ_z such that $\theta_x^T x = \theta_z^T Ax$ for all x. Finding such θ_z is equivalent to solving the overdetermined linear system $A^T \theta_z = \theta_x$, which can

be done only if the system is *consistent*, i.e. if it has solution. This will happen if and only if θ_x is in the span of the columns of A^T.
2

In that case, by looking at the equivalent system $AA^T\theta_z = A\theta_x$ we can identify two cases:

1. A has linearly independent rows. In this case AA^T is invertible, so there is a unique solution given by $\theta_z = (AA^T)^{-1}A\theta_x$.

2. A has linearly dependent rows. In this case, the system is indeterminate and has an infinite number of solutions.

The matrix $(AA^T)^{-1}A$ of part (i) is known as the Moore-Penrose pseudo-inverse of A^T, and it is denoted by $(A^T)^\dagger$.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

2. (d)

1/1 point (graded)
Given the same classifiers as in (b), if there is a θ_x that produces good classifications of the data vectors, will there **always** be a θ_z that will identically classify the associated z's?

Note: Now assume that you can change the $m \times n$ matrix A.

- [x] Yes ✔

No

Solution:

We now have flexibility in both A and θ_z. We want to find A, θ_z such that $A^\top \theta_z = \theta_x$. We can achieve this by simply setting $\theta_z = 1$, the first row of A to be θ_x, and the remaining rows to be 0:

$$A^\top \theta_z = \begin{bmatrix} | & | \\ \theta_x & 0 \\ | & | \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \theta_x$$

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

2. (e)

2/2 points (graded)

If $m < n$, can we find a more accurate classifier by training in z-space, as measured on the training data?

Yes

- (selected) No ✔
- Depends

How about on unseen data?

- Yes
- No
- (selected) Depends ✔

Solution:

- The accuracy in z-space is always bounded by the x space, as we can always construct a classifier in x space that corresponds to a classifier in z space.
- Without any assumption, the unseen data can be arbitrary. Hence, we can always construct a dataset that favors the classifier produced in z space. We can do the same thing to the classifier produced in x space as well.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Homework 3 / 2. Feature Vectors Transformation

Add a Post

Show all posts ▼ by recent activity ▼

2. (b)
I'm out in this problem. Are we expected to enter numbers of a formula? More over, If z is a [2x1] vector, does theta stand for a dot product facto...
9

2.e Seen and unseen data, supervised training.
It seems to me that the transformation from x space to z space take into account the label (supervised training) of the training data. Thus this tra...
2

A little hint for 2.a
I found that I was grossly overthinking this problem. It helped me to consider matrix-vector multiplication rules (https://www.varsitytutors.com/h...
7

[2.d] need explaining for solution
I understand if we can arbitrarily change matrix A, we can find a linear system which is always consistent. But in the example provided by solutio...
2

2d is confusing
Dear Staff, It is not clear why my answer is marked wrong. It is actually not clear to me that you are allowing me to change "m" in that question. I ...
3

Unfair scoring
All but 2e for me was right. I spend so much time to solve 2a – 2d, and all these tasks are worth only one point. I missed 2e and I lost 2 points. I d...
4

To be honest... This problem deserves more attempts
1 attempt does not clear the answers.
2

- Related lecture — 1 new_
  Hi. I remember when this assignment was first launched in HW 1, and a TA or a Staff warned us the topic was actually covered in lecture 5. I went...
- [Staff] Question 2(a) & 2(b) — 3
  Hi folks, I know there have been a number of discussions regarding 2(a). I have pretty much completed all other questions in this homework but ...
- just one attempt? — 2
  I think that one attempt in section 2e is not enough
- A is a fixed matrix. — 2
  Hello, I was wondering if the matrix A that you are referring to in the questions 2.c and 2.d could be any arbitrary one (let's say all zeros), but with...
- [staff] What can we tweak for 2d? — 6
  Community TA
- 2b - need to enter theta_{z} not working — 3