

Introduction to Machine Learning

Due: Monday, March 04, 2024 at 09:00 AM

For these exercises, it will be helpful to review the notes on Feature Representations and Classification.

1) Feature Representation

For each of the following features, pick what might be the best encoding for linear classification, among the choices provided. In each part below, you should assume that the dataset has several other features. The questions mention what we're trying to predict, but we're really just asking about the input encoding here.

The point of this question is to think about alternatives; there are several options, many not mentioned here.

1.1)

Car brand, e.g., Chevy, Ford, Toyota, VW, for predicting gas mileage (e.g., low vs. high gas mileage):

4 unary features (one-hot): 1000, 0100, 0010, 0001

100.00%

You have infinitely many submissions remaining.

1.2)

A person's weight (in kg) for predicting life expectancy (e.g., live for longer than 50 years vs. live for less than 50 years). Assume there is also a real-valued feature in the dataset encoding height in meters that we would like to use later on. A standardized real-valued feature

100.00%

You have infinitely many submissions remaining.

2) Feature Mapping

2.1)

Consider the following, one-dimensional, dataset:

$$\mathcal{D} = \{(-1, +1), (0, -1), (1, +1)\} .$$

It is not linearly separable in its original form.

Which of these feature transformations lead to a separable problem? (Mark all that apply)

- $\phi(x) = 0.5 * x$
- $\phi(x) = |x|$
- $\phi(x) = x^3$
- $\phi(x) = x^4$
- $\phi(x) = x^{2k}$ for any positive integer k
- $\phi(x) = x^{2k+1}$ for any positive integer k

100.00%

You have infinitely many submissions remaining.

2.2)

Your friend Kernelius uses feature transformation $\phi(x) = [x, x^2]^T$ on the data from the previous problem. In the new space, the linear classifier with $\theta = [0, 1]^T$ and $\theta_0 = -0.25$ achieves perfect accuracy. What points from the original space \mathbb{R} get classified as members of the negative class? (It may be helpful to find the equation of the separator.)

Enter a Python list denoting the inclusive range $[a, b]$ such that all values of $x \in [a, b]$ are classified to be members of the negative class (points on the separator are classified negative):

100.00%

You have infinitely many submissions remaining.

Survey

(The form below is to help us improve/calibrate for future assignments; submission is encouraged but not required. Thanks!)

How did you feel about the **length** of this exercise?

- Too long.
- About right.
- Too short.

How did you feel about the **difficulty** of this exercise?

- Too hard. We should tone it down.
- About right.
- Too easy. I want more challenge.

Do you have any feedback or comments about any questions in this exercise? Anything else you want us to know?