3. Introduction to Non-linear Classification

Exercises due Oct 14, 2020 05:29 IST *Completed*

Introduction to Non-linear Classification

Start of transcript. Skip to the end.

We can play this game further.

For any example x, let's say it's a scalar again,

we can add feature coordinates.

Not just a single one, but we can

take x, x to the power of 2, 3, 4, and so on.

We can always add additional feature

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Counting Dimensions of Feature Vectors

0/1 point (graded)

Let $x \in \mathbf{R}^{150}$, i.e. $x = \left[x_1, x_2, \ldots, x_{150}\right]^T$ where x_i is the i-th component of x. Let $\phi(x)$ be an **order 3** polynomial feature vector. This means, for example, $\phi(x)$ can be

$$\phi(x) = \left[\underbrace{x_1, \ldots, x_i, \ldots, x_{150}}_{\deg 1}, \underbrace{x_1^2, x_1 x_2, \ldots, x_i x_j, \ldots {x_{150}}^2}_{\deg 2}, \underbrace{{x_1}^3, {x_1}^2 x_2, \ldots, x_i x_j x_k, \ldots, {x_{150}}^3}_{\deg 3}\right] \quad \text{where } 1 \leq i$$

Note that the components of $\phi(x)$ forms a basis of the space of all polynomials with zero constant term and of degree at most 3.

What is the dimension of the space that $\phi(x)$ lives in? That is, $\phi(x) \in \mathbb{R}^d$ for what d?

Hint: The number of ways to select a multiset of k non-unique items from n total is $\binom{n+k-1}{k}$. For example, if a ball can be any of 3 colors, then the number of color configurations of 2 balls is $\binom{3+2-1}{2} = \binom{4}{2} = 6$.

$d =$ 3 ✖ **Answer:** 585275

Solution:

For each of the feature transformations (power 1, power 2, power 3), there are n-multichoose-power combinations. Thus $\binom{150}{1} + \binom{151}{2} + \binom{152}{3} = 585275$. **Remark:** We see that the dimension of the space that the feature vectors live grows quickly as a function of d, the dimension we started with if $x \in \mathbb{R}^d$.

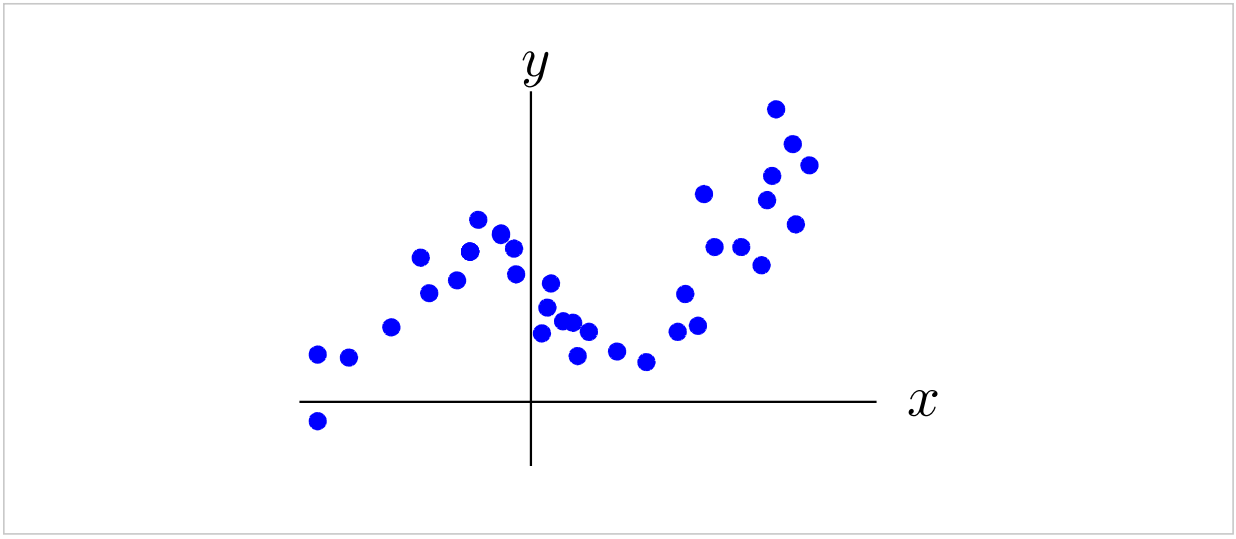
Submit You have used 3 of 3 attempts

Answers are displayed within the problem

Regression using Higher Order Polynomial feature

1/1 point (graded)

Assume we have n data points in the training set $\left\{ \left(x^{(t)}, y^{(t)}\right) \right\}_{t=1,\ldots,n}$ where $\left(x^{(t)}, y^{(t)}\right)$ is the t-th training example:



We want to find a non-linear regression function f that predicts y from x, given by

$$f\left(x; \theta, \theta_0\right) = \theta \cdot \phi\left(x\right) + \theta_0$$

where $\phi\left(x\right)$ is a polynomial feature vector of some order. What (loosely) is the minimum order of $\phi\left(x\right)$?

3 ✔ **Answer:** 3

Solution:

The relationship between y and x can be roughly described by a cubic function, so a feature vector $\phi\left(x\right)$ of minimum order 3 can minimize structural errors.

Submit You have used 1 of 2 attempts

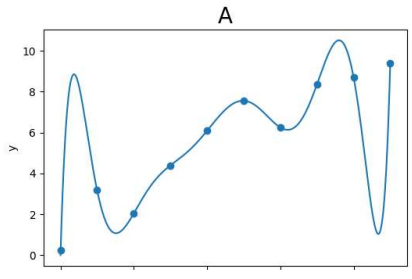
Answers are displayed within the problem

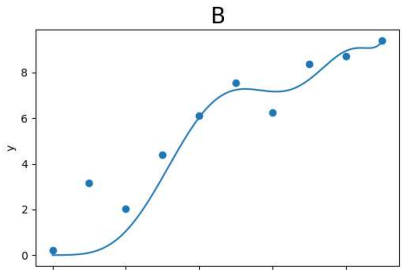
Effect of Regularization on Higher Order Regression

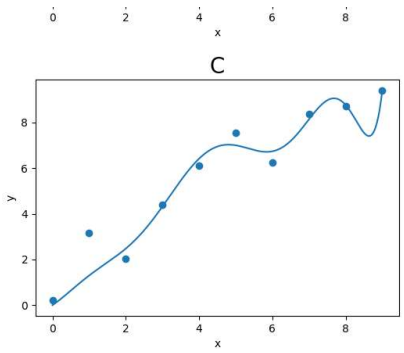
2/2 points (graded)

Let us go back to explore the effect of regularizaion on Higher Order regression.

The three figures below show the fitting result of a 9th order polynomial regression with different regularization parameter λ on the same training data.







Which figure above corresponds to the smallest regularization parameter λ?

- (●) A
- ( ) B
- ( ) C

✔

Which figure corresponds to the largest regularization parameter λ?

- ( ) A
- (●) B
- ( ) C

✔

Solution:

The effect of regularization is to restrict the parameters of a model to freely take on large values. This will make the model function smoother, leveling the 'hills' and filling the 'vallyes'. It will also make the model more stable, as a small perturbation on x will not change y significantly with smaller $\|\theta\|$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 6. Nonlinear Classification / 3. Introduction to Non-linear Classification

Add a Post

Show all posts by recent activity

- [Staff] About Q.2 Answer — 3
  First I've done question 2 correctly but than mistakenly i wrote answer of Q.1 to Q.2 block and now its says question failed. Can staff hel...
- How can we get minimum order just by looking at graph — 6
  How can we get minimum order just by looking at graph
- Incorrect video transcripts. — 1
  The video transcripts are really sloppy. I generally do not use them. But when I am actually not sure what is being told I look it up and it ...
- hint for Q2 — 3
  I find the explanation is nice. http://openclassroom.stanford.edu/MainFolder/DocumentPage.php?course=MachineLearning&doc=exerci

I find the explanation is nice: http://openclassroom.stanford.edu/MainFolder/DocumentPage.php?course=MachineLearning&doc=exerci...

	Got # dimensions correct, but is it really?	5
?	Counting Dimensions of Feature Vectors Hi, I have seen the solution for the question " Counting Dimensions of Feature Vectors " . But i dont get it... How the compute work ? In t...	11
?	How to count The number of ways to select a multiset of k non-unique items from n total? I'm doing The MIT probability course in parallel but it doesn't cover this counting technique. Any suggestions on that?	7

‹ Previous | Next ›