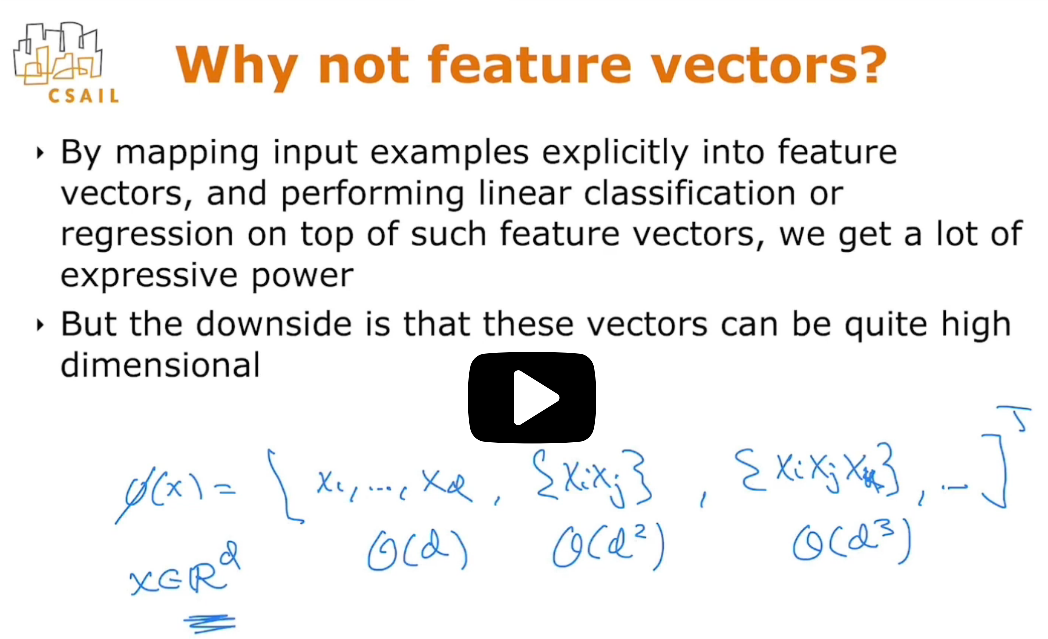
# 3. Introduction to Non-linear Classification

## Introduction to Non-linear Classification




8:18 / 8:18 Speed 1.50x

So our feature vector becomes very high-dimensional

very quickly if we even started from a moderately dimensional

vector.

OK?

So we would want to have a more efficient way of doing that--

operating with high dimensional feature

vectors without explicitly having to construct them.

**And that is what kernel methods provide us.**

### Video

Download video file

### Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

## Counting Dimensions of Feature Vectors

1/1 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

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

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 $\binom{3+2-1}{2} = \binom{4}{2} = 6$.

$d =$ 585275 ✔ **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 $x \in \mathbb{R}^d$.

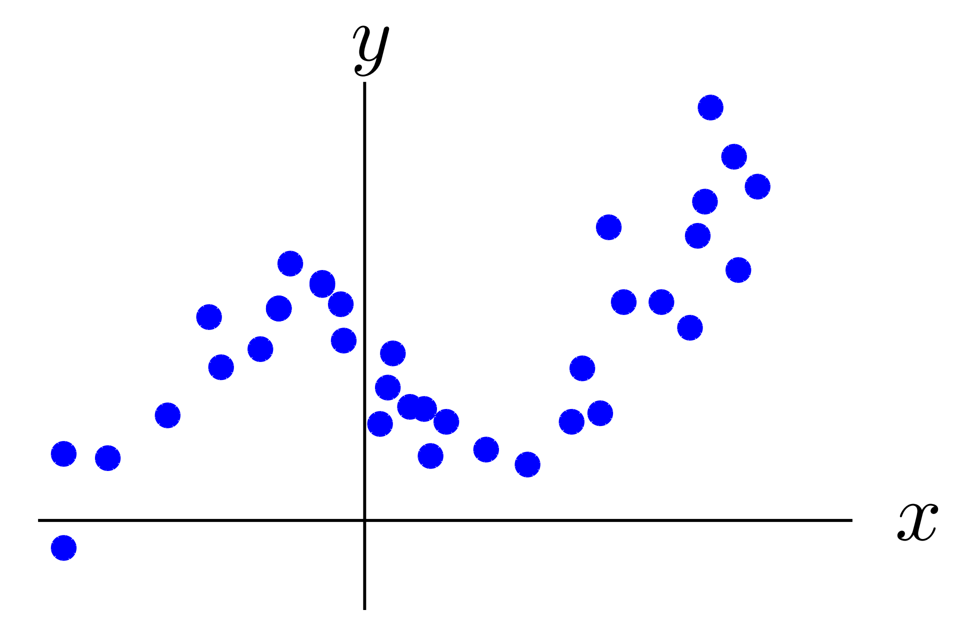
Submit You have used 1 of 3 attempts

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 $\left\{ (x^{(t)}, y^{(t)}) \right\}_{t=1,\ldots,n}$ where $(x^{(t)}, y^{(t)})$ 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

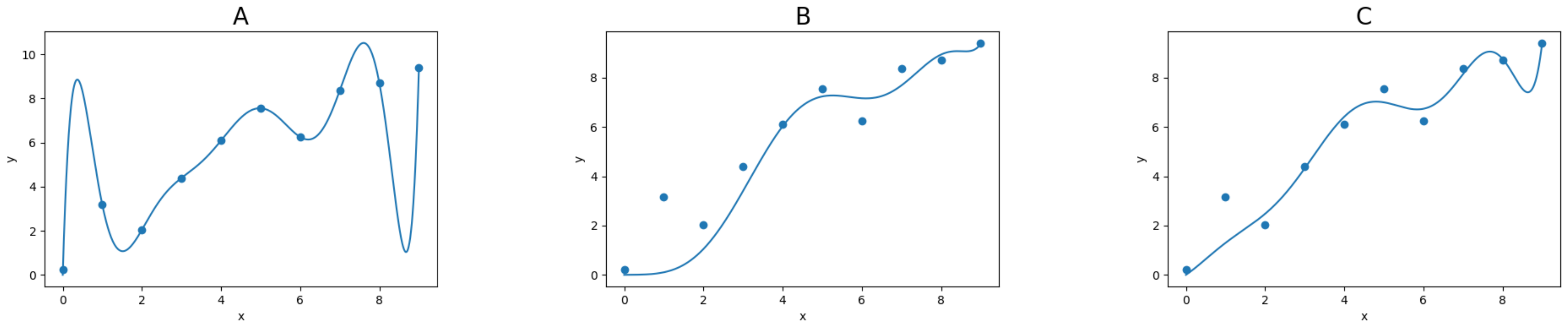
Answers are displayed within the problem

## 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 an 8th order polynomial regression with different regularization parameter $\lambda$ on the same training data.




Which figure above corresponds to the smallest regularization parameter $\lambda$?

- A ✔
- B
- C

Which figure corresponds to the largest regularization parameter $\lambda$?

- 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

All Posts

### Counting Dimensions of Feature Vectors

question posted 5 days ago by michael_x

1) Once $x_1x_2$ appeared in the sequence, $x_2x_1$ won't appear ?

2) "if a ball can be any of 3 colors, then the number of color configurations of 2 balls is" $3 \times 3 = 9$, isn't it ?

This post is visible to everyone.

Add a Response

6 responses

**Mark_B2** (Community TA)
5 days ago

1) are they any different?

Add a comment

**nbourbon**
5 days ago

they are the same, so we should assume that those combinations shouldn't be counted right?

Correct

posted 5 days ago by **robweatherston18**

Add a comment

**alihuengarciapav**
4 days ago

1) Yes. They are the same number, it wouldn't have any sense to consider them different.

2) No. And this involves 1). If you have 9 possibilities is because you are considering AB and BA as different options, and they are not.

Don't forget that this technic pretends to make a polynomial with the original feature vector. Then, if you have, for example, $x_1x_2 + x_2x_1$, this is just $2x_1x_2$.

Please correct me if someone sees that I'm wrong.

I got the correct answer just by a kind of brute force.

But I appreciate your post which tries to connect my questions 1) and 2).

By the way, do you agree with the following statement ?

"If a ball can be any of 3 colors, then the number of color configurations of 2 balls is" $3 \times 3 = 9$ is true if we don't have any further contexts.

posted 4 days ago by **michael_x**

They are implying that if we got : R, G, B balls, then RG = GB and should be counted as one. To be frank i did not know/remember the formula in the hint, i got it by breaking the each deg into pieces and applying unique combinations in a set, meaning number of subset of specific size.

posted 4 days ago by **RommelAlbertoRodriguezPerez**

This is analogous to the number of ways in which we can place $k$ identical balls into $n$ numbered boxes. We can put each allocation of balls to boxes of this problem into a 1-to-1 correspondence with lining up $k$ balls and $n-1$ (identical) dividers in a particular order. E.g. $||00|0||0$ corresponds to placing $2$ balls in box $3$, $1$ ball in box $4$ and $1$ ball in box $6$. It's easy to see the number of those arrangement is $\binom{k+n-1}{k} = \binom{k+n-1}{n-1}$. Therefore no need to remember the formula, but only this "story proof" (I learned this term from Joe Blitzstein in his great Stat 110 course.)

posted 2 days ago by **mrBB** (Community TA)

Add a comment

**nr7116**
4 days ago

order 1 :n = 2, k = 1 : $[x_1, x_2]$

order 2 : n = 2, k = 2 : $[x_1^2, 2x_1x_2, x_2^2]$

order 3 :n = 2, k = 3 : $[x_1^3, 3x_1^2x_2, 3x_1x_2^2, x_2^3]$

Add a comment

**kunapalli**
4 days ago

Sets do not have an order. Therefore AB same as BA in the color configurations. We use sets in this context because in arithmetic a*b = b*a.

Add a comment

**mrBB** (Community TA)
2 days ago

$x_1x_2$ and $x_2x_1$ can never be both in a basis of our space of functions because they trivially and obviously aren't independent.

Add a comment

Showing all responses

Add a response:

Preview

Submit

Learn About Verified Certificates