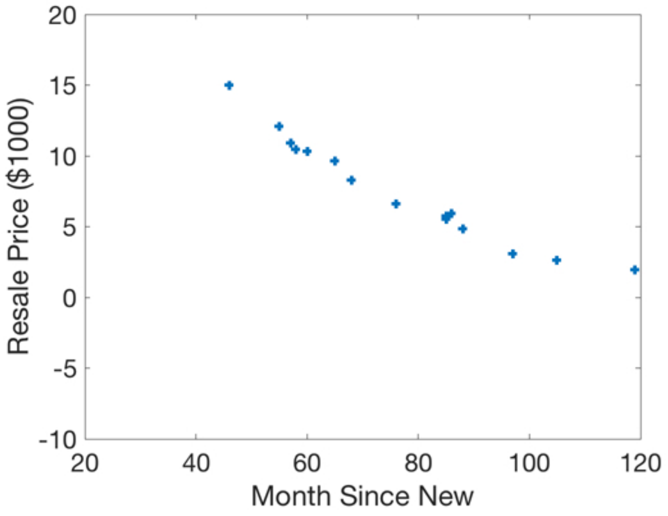
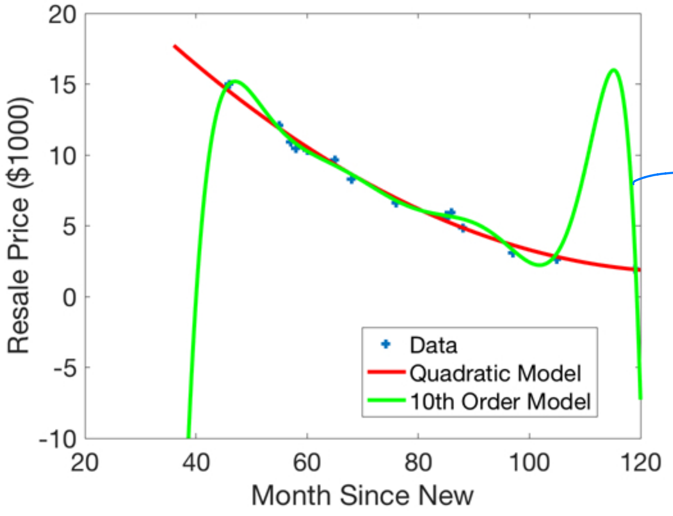
Patterns in Data and Outer Products

Objectives

- introduce low-dimensional modeling
- define <u>outer product</u>
- use outer products to model matrices
- introduce taste profiles for ratings

Patterns and Model Order





prediction of price → time

$$\hat{p} = f(t)$$ polynomial

$$\hat{p}_i = \begin{bmatrix} 1 & t_i & t_i^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$ Quadratic

"label" "feature" "model"

Another way to express

$$\begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \vdots \\ \hat{p}_{20} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{20} & t_{20}^2 \end{bmatrix}}_{T} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} w_0 + \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_{20} \end{bmatrix} w_1 + \begin{bmatrix} t_1^2 \\ t_2^2 \\ \vdots \\ t_{20}^2 \end{bmatrix} w_2$$

building blocks - bases

$$\hat{p} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} w_0 + \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_{20} \end{bmatrix} w_1 + \cdots + \begin{bmatrix} t_1^{10} \\ t_2^{10} \\ \vdots \\ t_{20}^{10} \end{bmatrix} w_{10}$$

Data Fit ✓

Higher Dimension vs

✗ Generalization

Comments

- computing $\underline{I}\,\underline{w}$ involves inner products with rows of $\underline{I}$
- interpreting model with "bases" uses columns of $\underline{I}$
- model order trades data fit and generalization errors

The power of the function

(☆ Express as weighted sums of each feature column)

Modeling Matrix Data

$$\underline{R} = \begin{array}{c|c} & \begin{array}{cccc} \text{Noah} & \text{Emily} & \text{Emma} & \text{Liam} \end{array} \\ & \begin{bmatrix} -2 & 3 & -1 & 4 \\ -4 & 3 & -1 & 2 \\ 4 & -2 & 0 & ? \\ 2 & -4 & 2 & -3 \end{bmatrix} \end{array} \quad \begin{array}{l} \text{movie} \\ \text{Last Jedi} \\ \text{Star Trek} \\ \text{When Harry Met Sally} \\ \text{Ground Hog Day} \end{array}$$

ratings -5 to 5

Use a "taste profile" to model each user's preferences

$$\underline{t}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \begin{array}{l} \}\text{sci fi} \\ \\ \}\text{rom com} \end{array}$$

Noah: $\sim -3\underline{t}_1 = \begin{bmatrix} -3 \\ -3 \\ 3 \\ 3 \end{bmatrix}$

Emily: $\sim 3\underline{t}_1 = \begin{bmatrix} 3 \\ 3 \\ -3 \\ -3 \end{bmatrix}$

$$\hat{\underline{R}} = \begin{bmatrix} -3 & 3 & -1 & 3 \\ -3 & 3 & -1 & 3 \\ 3 & -3 & 1 & -3 \\ 3 & -3 & 1 & -3 \end{bmatrix}$$

☆ Predefined data across general audience → For approximate individual user's taste

$\hat{\underline{R}}$ can be expressed as an "outer product"

$$\hat{\underline{R}} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \underbrace{\begin{bmatrix} -3 & 3 & -1 & 3 \end{bmatrix}}_{\underline{s}_1^T} = \underline{t}_1 \underline{s}_1^T = \begin{bmatrix} -3 & 3 & -1 & 3 \\ -3 & 3 & -1 & 3 \\ 3 & -3 & 1 & -3 \\ 3 & -3 & 1 & -3 \end{bmatrix}$$

★ Scaling factor from the last page (Each value represents a user's custom factor)

column x row = matrix

Include another taste profile

$$\underline{t}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

Carrie Fisher movies

★ More than one axis of variation in preferences

$$\hat{\underline{R}} = \underline{t}_1 \underline{s}_1^T + \underline{t}_2 \underline{s}_2^T$$

$$\underline{s}_2^T = \begin{bmatrix} 1 & 1/2 & -1/2 & 5/3 \end{bmatrix}$$

$$\hat{\underline{R}} = \underbrace{\begin{bmatrix} -3 & 3 & -1 & 3 \\ -3 & 3 & -1 & 3 \\ 3 & -3 & 1 & -3 \\ 3 & -3 & 1 & -3 \end{bmatrix}}_{\text{Taste 1}} + \underbrace{\begin{bmatrix} 1 & 1/2 & -1/2 & 5/3 \\ -1 & -1/2 & 1/2 & -5/3 \\ 1 & 1/2 & -1/2 & 5/3 \\ -1 & -1/2 & 1/2 & -5/3 \end{bmatrix}}_{\text{Taste 2}} = \underbrace{\begin{bmatrix} -2 & 3.5 & -1.5 & 4.7 \\ -4 & 2.5 & -0.5 & 1.3 \\ 4 & -2.5 & 0.5 & -1.3 \\ 2 & -3.5 & 1.5 & -4.7 \end{bmatrix}}_{\text{More refined predictions}}$$

Interpret matrix multiplication as sum of outer products

$$\hat{\underline{R}} = \left[\, \underline{t}_1 \,\vdots\, \underline{t}_2 \,\right] \begin{bmatrix} \underline{s}_1^T \\ \hdashline \underline{s}_2^T \end{bmatrix} = \underline{t}_1 \underline{s}_1^T + \underline{t}_2 \underline{s}_2^T$$

In general,

$$\underset{N \times K}{\underline{R}} = \underset{N \times L}{\underline{T}}\ \underset{L \times K}{\underline{S}} = \underset{N \times L}{\left[\, \underline{t}_1 \,\vdots\, \underline{t}_2 \,\vdots\, \cdots \,\vdots\, \underline{t}_L \,\right]} \underset{L \times K}{\begin{bmatrix} \underline{s}_1^T \\ \hdashline \underline{s}_2^T \\ \hdashline \vdots \\ \hdashline \underline{s}_L^T \end{bmatrix}} = \sum_{\ell=1}^{L} \underbrace{\underline{t}_\ell \underline{s}_\ell^T}_{N \times K}$$

- Choose L to trade model fit and generalization
- Methods for finding good $\underline{T}$, $\underline{S}$, L discussed later
- Inner products for computation, outer products interpretation