机器学习和量化交易实战

第二讲

这次和下次课程的任务目标

第二节课和第三节课是一个小单元，主要包括如下内容：

本次课：

1. 掌握python语言和常用数据处理包
2. 从技术分析到机器学习

下次课（你们要的数据和程序，finally)

1. 实战：python爬取金融数据
2. 实战2: 利用python进行金融数据处理：数据清洗，数据可视化，特征提取，etc.
3. 实战3:你的第一个基于机器学习的量化模型（yay）

需要掌握的python的知识点

主要平台：

Anaconda的安装

ipython notebook

需要掌握的python的知识点

1. Python 的数据类型

    str,float,bool,int,long

1. python的基本语法：分支，循环，函数
2. python的数据结构：tuple,list,dictionary,etc
3. python的内置函数
4. python和面向对象编程

自学地址： https://learnxinyminutes.com/docs/python/

需要掌握的numpy的知识点

1. 利用numpy进行各类线性代数的运算：
    1. 创建矩阵，向量，etc
    2. 熟练掌握矩阵的索引

2. numpy的输入和输出
3. numpy的常用函数

自学地址：书籍《利用python进行数据分析》第四章

需要掌握的pandas的知识点

1. pandas与数据io

2. pandas 的dataframe的各种内置函数（统计指标，绘图）

3. pandas的索引

自学地址：书籍 《利用python进行数据分析》第5章

需要掌握的sklearn的知识点

1. 利用sklearn在mnist数据上做分类
2. 利用sklearn做线性回归模型

http://scikit-learn.org/stable/auto_examples/index.html

这只股票要不要买

账面价值：
- 10 * 10万 工厂
- 专利 100万
- 20万负债

内在价值
- 1 万 分红 / 年 5%的折现率

市场价值
- 1万股
- 每股75块钱

这只股票要不要买

账面价值：80万

- 10 * 100万 工厂
- 专利 100万
- 20万负债

内在价值 20万

- 1 万 分红 / 年 5%的折现率

市场价值 75万

- 1万股
- 每股75块钱

CAPM Model

Portfolio 资产组合

［a%, b%, c%]

abs（a%）+abs(b%)+ abs(c%) = 100%

Market Portfolio

SP500

沪深三百

Etc

个股的CAPM model

$r_i(t) = beta_i * r_m(t) + alpha_i(t)$

CAPM says

E(alpha(t)) = 0

Linear scaled return of the market, with some noise at mean 0.

被动式管理 vs 主动式管理基金

被动式管理：复制大盘指数，持有。

主动式管理：选择个股，频繁交易

$$r_i(t) = beta_i * r_m(t) + alpha_i(t)$$

关键分歧：

Alpha 是否是随机噪声， alpha的期望值是否为零。

投资组合的CAPM 模型

$$r_p(t) = \sum_i w_i\left(\beta_i r_m(t) + \alpha_i(t)\right)$$

$$= \sum_i \left[w_i \beta_i r_m(t) + w_i \alpha_i(t)\right]$$

$$= \underline{\sum w_i \beta_i}\, r_m(t) + \sum w_i \alpha_i(t)$$

$$r_p(t) = \beta_p r_m(t) + \alpha_p(t)$$

几个推论

E（alpha）＝0

选择好的beta值。

牛市：大beta

熊市：小beta

如果市场有效假说成立，我们无法预测股市，也选不出来合适的beta

价格套利理论（APT）

$r_i(t) = beta_i * r_m(t) + alpha_i(t)$

Beta 不是常数，而是一个变量。

Beta ＝ w＊r

两只股票的例子

Stock A: +1% mkt , beta = 1.0

Stock B: -1% mkt , beta_b = 2.0

Long A, short B.

技术分析 vs 基本面分析

历史数据：

- 价格，交易量
- 计算指标（features）
- 启发式选择（经验，机器学习）

技术分析何时works？

多个指标的非线性组合（机器学习）

短时

异类监测

最基本的指标以及机器学习怎么介入

Momentum 动量线 mom[t] = price[t] / (price[t-n]) – 1

SMA : Simple Moving Average. (smooth, laggged) … 可以看作一种滤波器。

BB （bollinger bands） BOLL指标： 决策边界是两个标准差

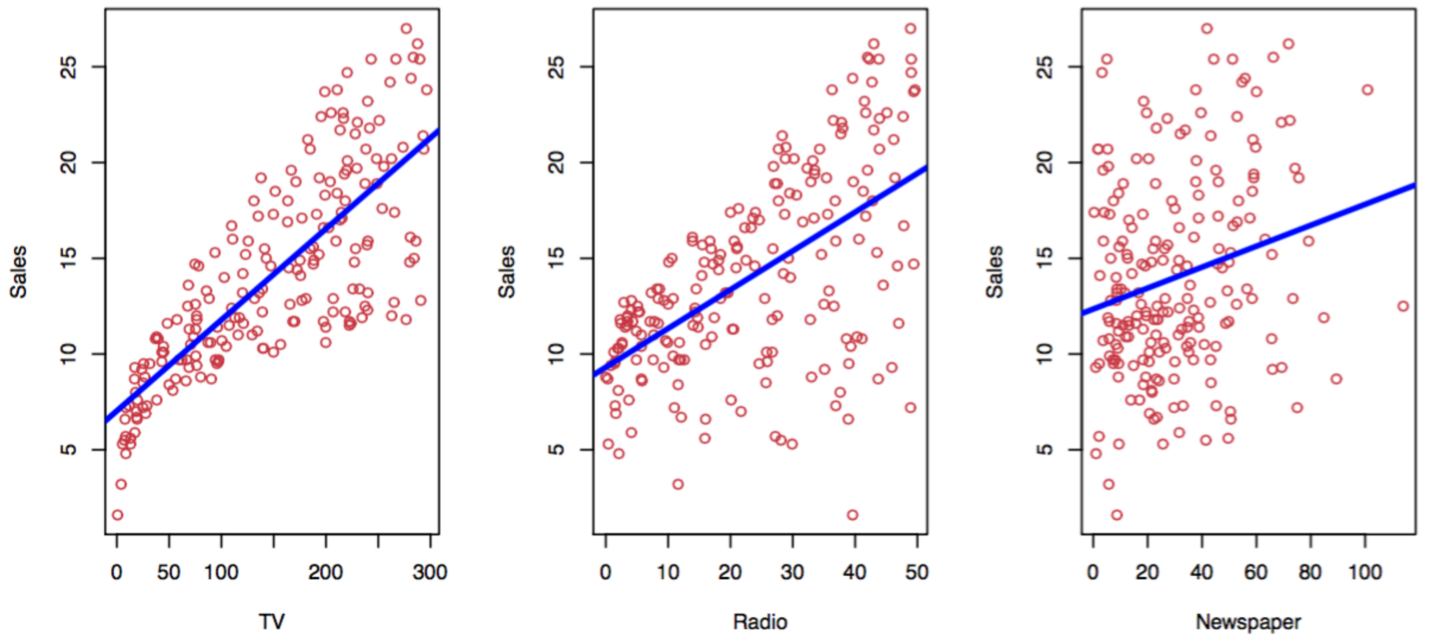
Normalization

SMA －0.5 ＋0.5

Mom －0.5， ＋0.5

BB －1， ＋1

Norm = (value – mean)/values.std()



Shown are Sales vs TV, Radio and Newspaper, with a blue linear-regression line fit separately to each.
Can we predict Sales using these three?
Perhaps we can do better using a model

$$\texttt{Sales} \approx f(\texttt{TV}, \texttt{Radio}, \texttt{Newspaper})$$

Here `Sales` is a *response* or *target* that we wish to predict. We generically refer to the response as Y.
`TV` is a *feature*, or *input*, or *predictor*; we name it X_1.
Likewise name `Radio` as X_2, and so on.
We can refer to the *input vector* collectively as

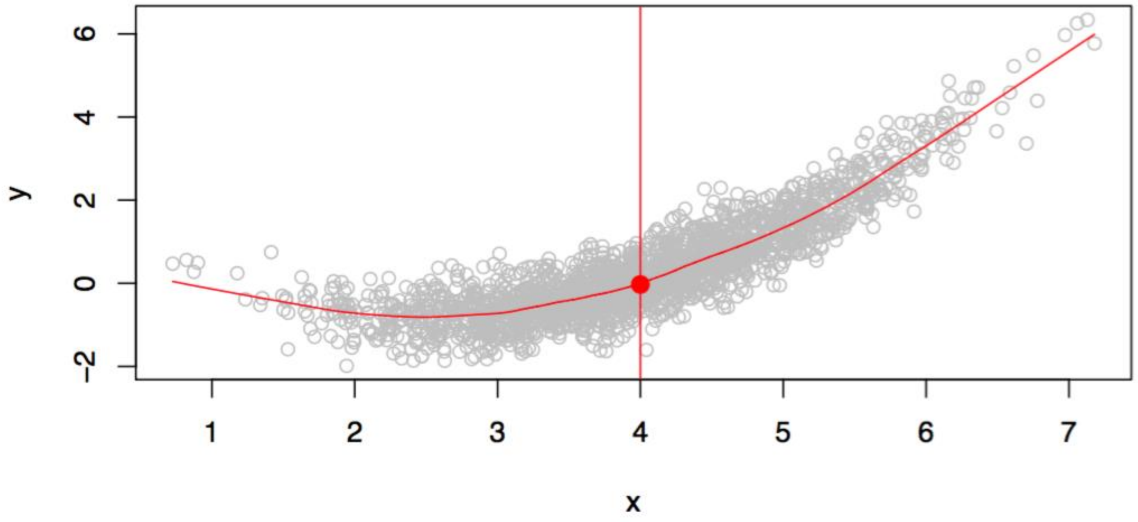
$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

Now we write our model as

$$Y = f(X) + \epsilon$$

where ϵ captures measurement errors and other discrepancies.

- With a good f we can make predictions of Y at new points $X = x$.
- We can understand which components of $X = (X_1, X_2, \ldots, X_p)$ are important in explaining Y, and which are irrelevant. e.g. `Seniority` and `Years of Education` have a big impact on `Income`, but `Marital Status` typically does not.
- Depending on the complexity of f, we may be able to understand how each component X_j of X affects Y.



Is there an ideal $f(X)$? In particular, what is a good value for $f(X)$ at any selected value of X, say $X = 4$? There can be many Y values at $X = 4$. A good value is

$$f(4) = E(Y|X = 4)$$

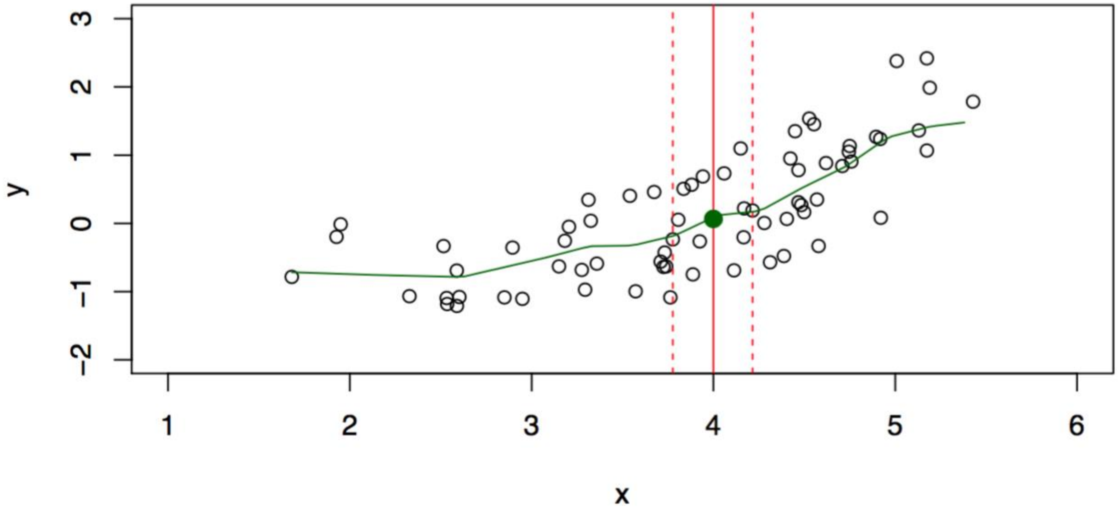
$E(Y|X = 4)$ means *expected value* (average) of Y given $X = 4$.

This ideal $f(x) = E(Y|X = x)$ is called the *regression function*.

- Typically we have few if any data points with $X = 4$ exactly.
- So we cannot compute $E(Y|X = x)$!
- Relax the definition and let

$$\hat{f}(x) = \text{Ave}(Y|X \in \mathcal{N}(x))$$

where $\mathcal{N}(x)$ is some *neighborhood* of x.

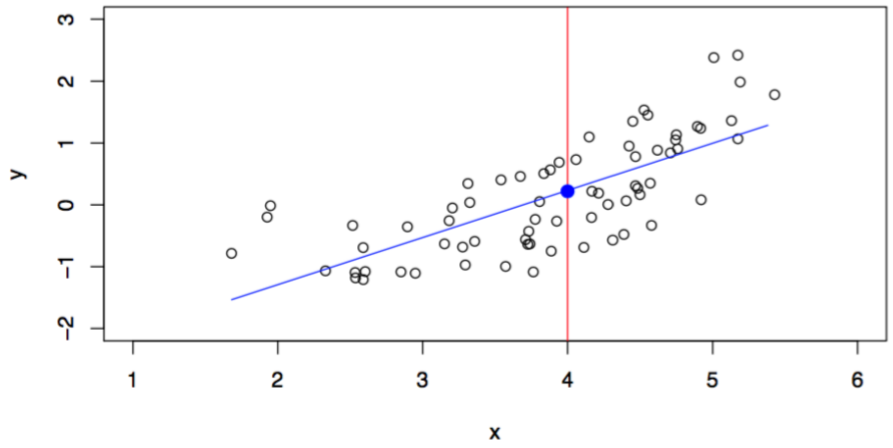


The *linear* model is an important example of a parametric model:

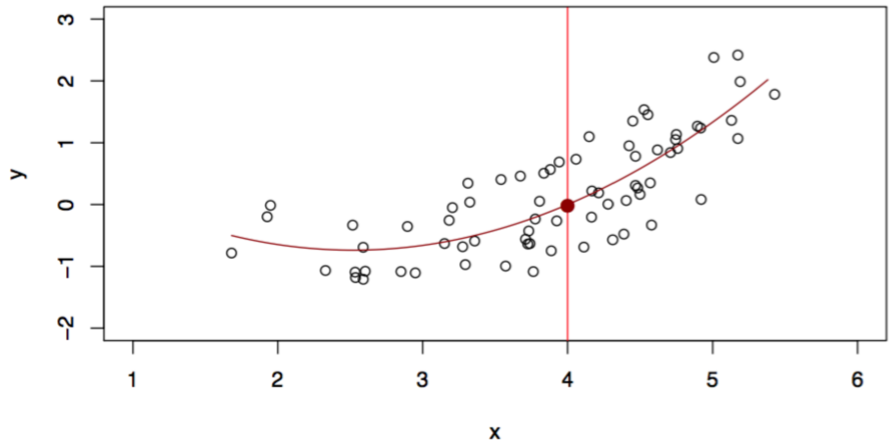
$$f_L(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_p X_p.$$

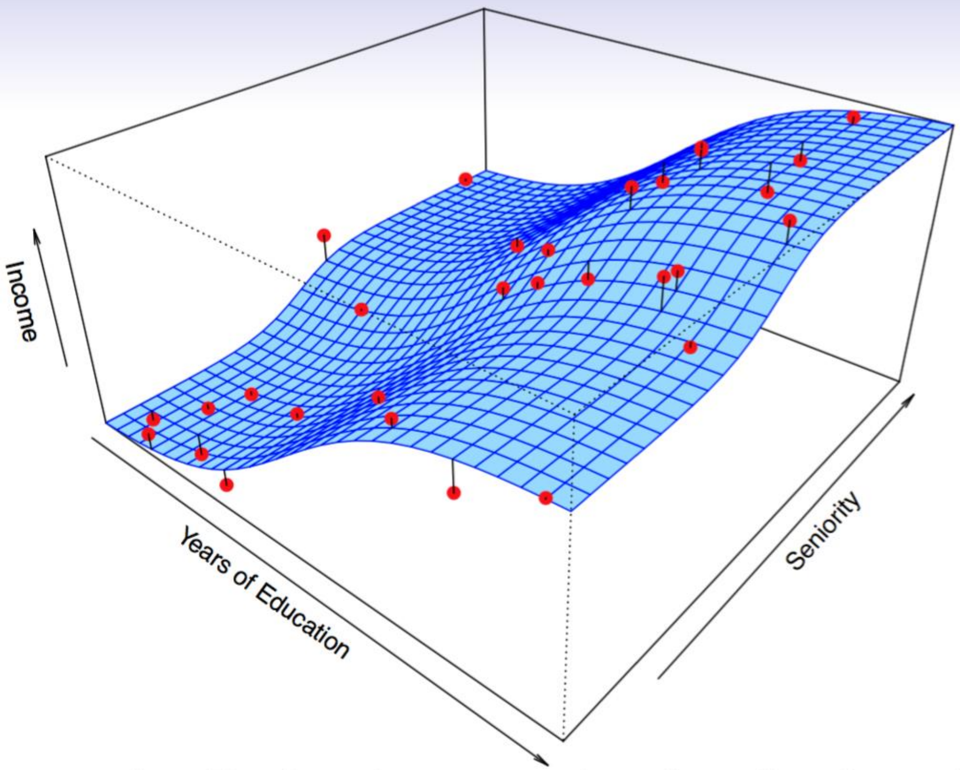
- A linear model is specified in terms of $p + 1$ parameters $\beta_0, \beta_1, \ldots, \beta_p$.
- We estimate the parameters by fitting the model to training data.
- Although it is *almost never correct*, a linear model often serves as a good and interpretable approximation to the unknown true function $f(X)$.

A linear model $\hat{f}_L(X) = \hat{\beta}_0 + \hat{\beta}_1 X$ gives a reasonable fit here



A quadratic model $\hat{f}_Q(X) = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2$ fits slightly better.

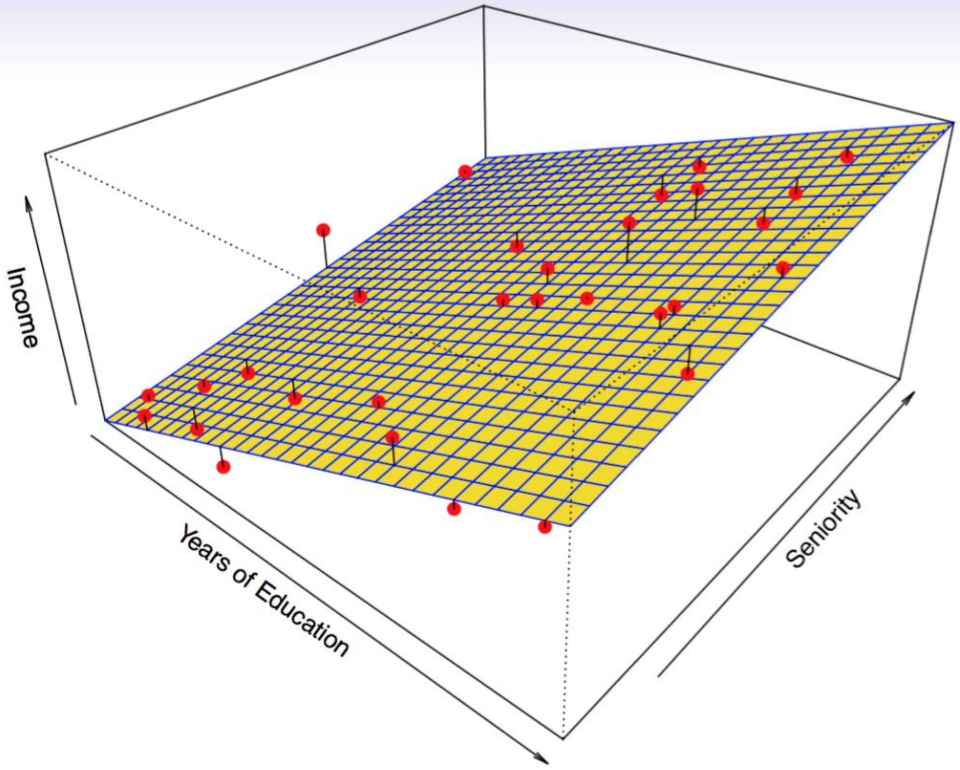




Simulated example. Red points are simulated values for $\texttt{income}$ from the model

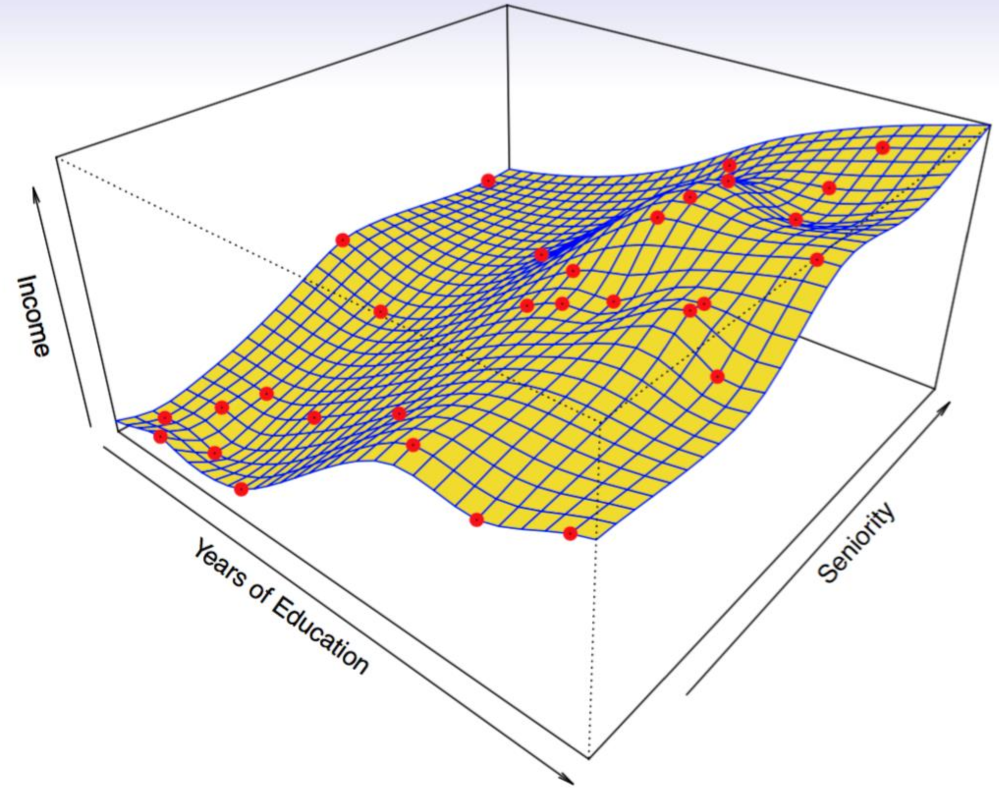
$$\texttt{income} = f(\texttt{education}, \texttt{seniority}) + \epsilon$$

f is the blue surface.

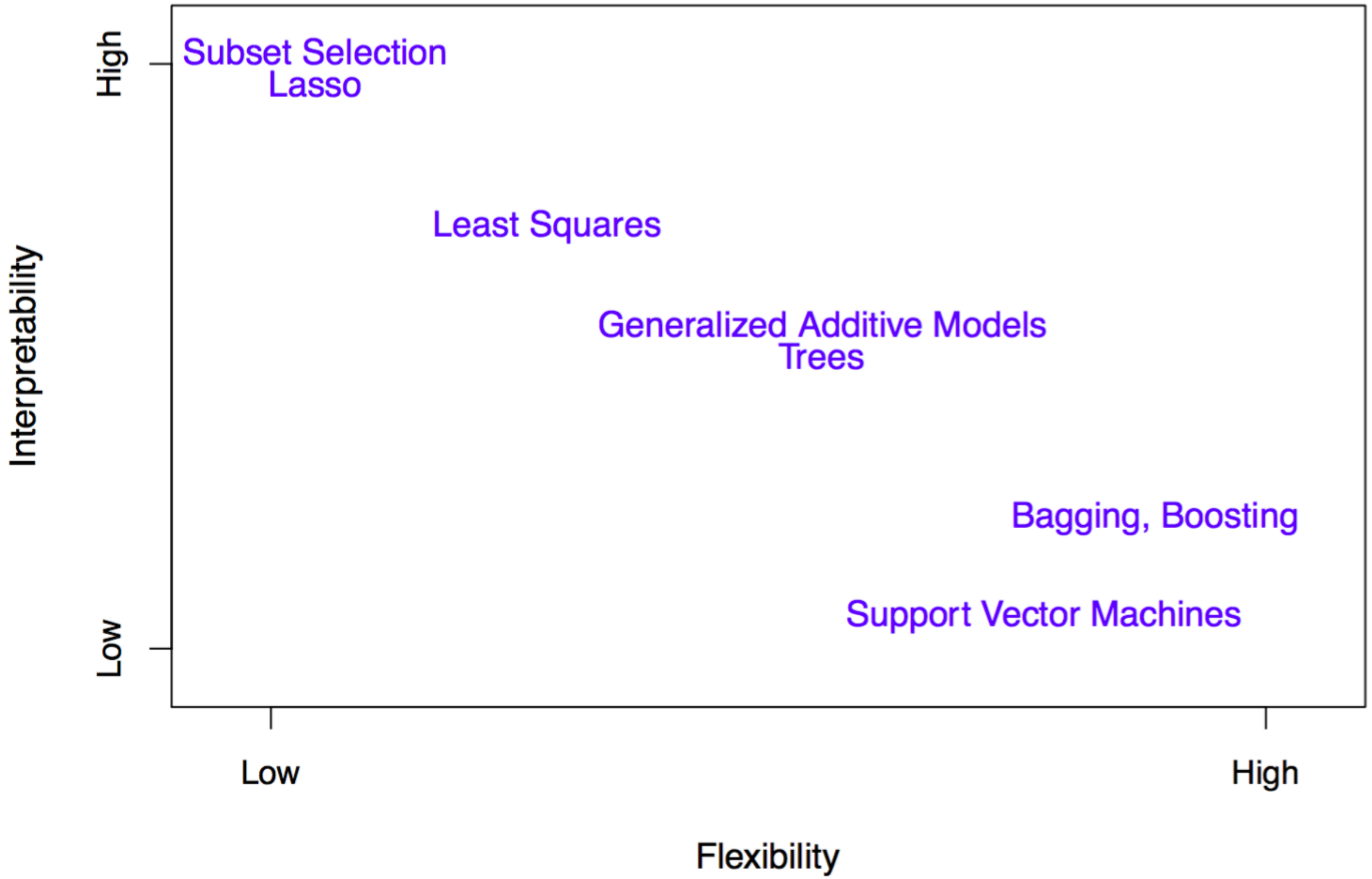


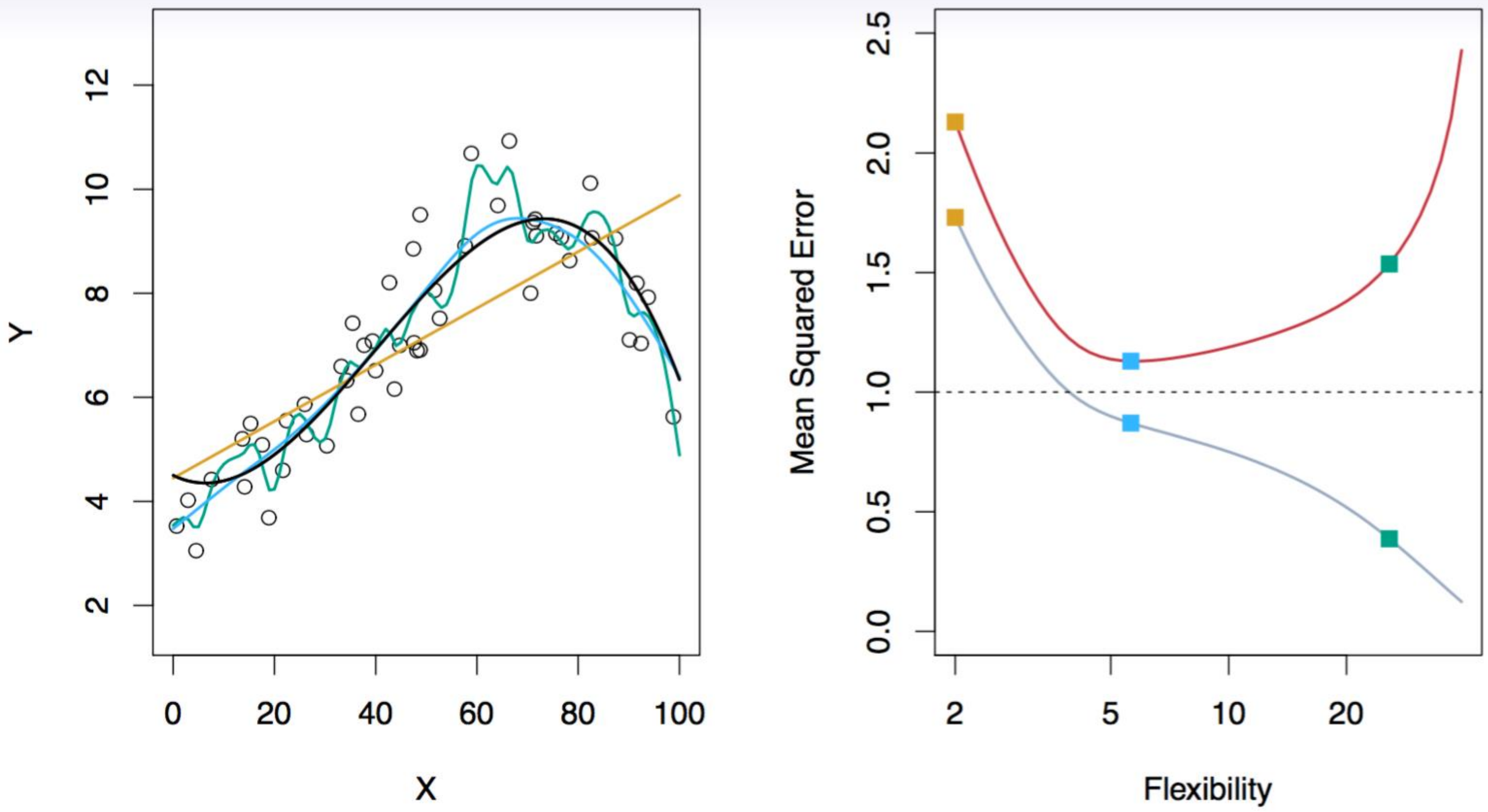
Linear regression model fit to the simulated data.

$$\hat{f}_L(\texttt{education}, \texttt{seniority}) = \hat{\beta}_0 + \hat{\beta}_1 \times \texttt{education} + \hat{\beta}_2 \times \texttt{seniority}$$



Even more flexible spline regression model $\hat{f}_S$(**education**, **seniority**) fit to the simulated data. Here the fitted model makes no errors on the training data! Also known as *overfitting*.


Subset Selection
Lasso
Least Squares
Generalized Additive Models
Trees
Bagging, Boosting
Support Vector Machines
High
Low
Interpretability
Low
High
Flexibility



Black curve is truth. Red curve on right is MSE_{Te}, grey curve is MSE_{Tr}. Orange, blue and green curves/squares correspond to fits of different flexibility.

Suppose we have fit a model $\hat{f}(x)$ to some training data Tr, and let (x_0, y_0) be a test observation drawn from the population. If the true model is $Y = f(X) + \epsilon$ (with $f(x) = E(Y|X = x)$), then

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \text{Var}(\hat{f}(x_0)) + [\text{Bias}(\hat{f}(x_0))]^2 + \text{Var}(\epsilon).$$

The expectation averages over the variability of y_0 as well as the variability in Tr. Note that $\text{Bias}(\hat{f}(x_0))] = E[\hat{f}(x_0)] - f(x_0)$.

Typically as the *flexibility* of $\hat{f}$ increases, its variance increases, and its bias decreases. So choosing the flexibility based on average test error amounts to a *bias-variance trade-off.*

Homework

掌握上述知识，我们下节课要上机了