Johns Hopkins Engineering

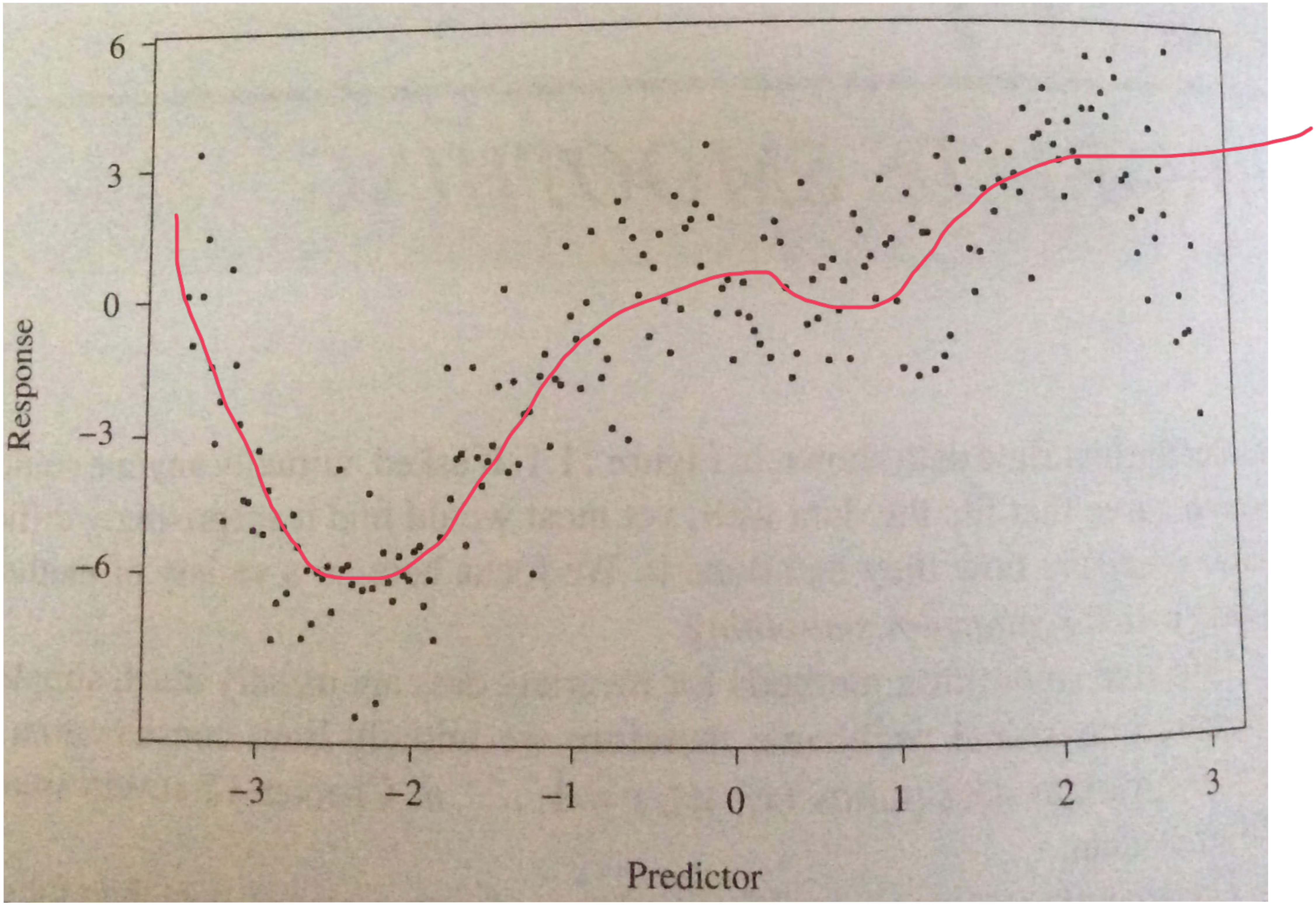
625.464 Computational Statistics

Introduction to Bivariate Smoothers

Module 12 Lecture 12A

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Introduction to Bivariate Smoothing



Bivariate Smoothing

Suppose we have n bivariate data points

$(X_i, Y_i)\quad i = 1, \dots, n$

that are <u>predictor</u>-<u>response</u> data

ie Y is assumed to be a (stochastic) function of X

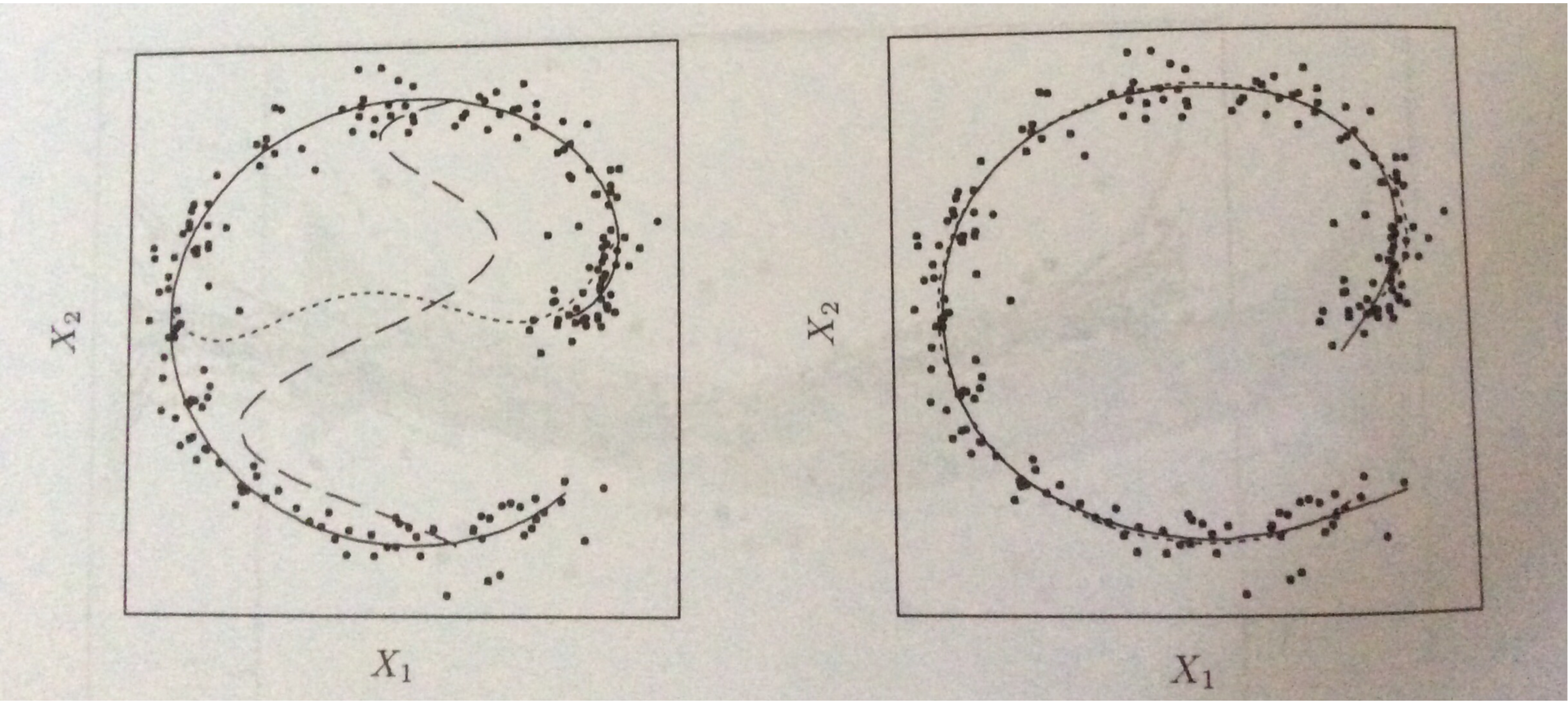
Then cond dist $Y|X$ describes how Y depends on $X=x$ and is a possible choice for smoother.

Ex/ $$Y_i = S(X_i) + \varepsilon_i$$

smooth function

stochastic noise mean 0

Why we use Predictor Response Dats



$$(x(t), y(t)) = ((1-\cos t)\cdot\cos(t), (1-\cos(t))\cdot\sin(t))$$

Predictor Response Data

Suppose we have P-R data (x_i, y_i) and $E[Y|X] = S(X)$ for a smooth function of X, S.

The Goal: Estimate $S(\cdot)$

smoother $\hat{S}(x)$.

→ based on data (x_i, y_i) and also a user specified smoothing parametr λ.

$\hat{S}_\lambda(X)$

Two Metrics for Predictor Response Data

For a given point x, let an estimate of $s(x)$ be $\hat{s}_\lambda(x)$. How do we know if $\hat{s}_\lambda(x)$ is good?

① $MSE_\lambda(\hat{S}_\lambda(x)) = E[(\hat{S}_\lambda(x) - s(x))^2] =$

$= (\text{bias}(\hat{S}_\lambda(x)))^2 - \text{var}(\hat{S}_\lambda(x))$ pointwise

consider a new point x^* for which we predict $\hat{s}_\lambda(x^*)$

② $MSPE_\lambda(\hat{S}_\lambda(x)) = E[(Y - \hat{S}_\lambda(x^*))^2 \mid x = x^*]$

$= \text{var}[Y \mid x = x^*] + MSE_\lambda(\hat{S}_\lambda(x^*))$ pointwise

How do we construct good smoothers?

Basic idea: Want the smoother to summarize the cond dist of Y_i given $X_i = x_i$ by some measure of location "center".

- want $\hat{s}(x)$ to go through the "center"
- in general they rely on local averaging

smooth at $x \to$ use Y_i whose x_i are close to x.

neighborhood of x

Generically: $\hat{S}_\lambda(x) = \text{ave}\{Y_i \mid x_i \in N(x)\}$

Comments on Lambda

Generic smoother

$$\hat{S}_\lambda(x) = \text{ave}\{y_i \mid x_i \in N(x)\}$$

λ will represent the span of the neighborhood $N(x)$

small $\lambda \rightarrow$ local $\rightarrow$ higher variance

large $\lambda \rightarrow$ distant points $\rightarrow$ introduce bias