STAT 391

2/13/2025

Lecture 12

KDE

Bias-Variance

Sol 3
HW4 due
HW5 out TB
lll.pdf

Lecture Notes V – Non-parametric density estimation

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

January, 2025

Kernels ✓

Kernel density estimators ✓ (Recap.)

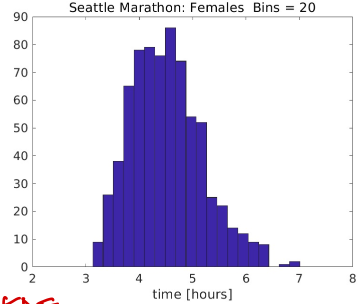
Train, Test, model class, model

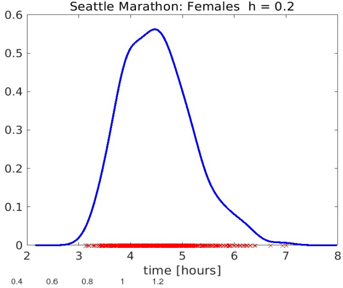
width

Choosing h by Cross-Validation and the Bias-Variance trade-off

Reading: Ch. 7

Kernel density estimators
for marathon data





KDE

Given $\mathcal{D} = x^{1:n} \in (-\infty, \infty)$ ~iid from P unknown

Model class : k, h = width ($h = \sigma$ for Normal kernel)

Want KDE of P (or PDF of P)

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x - x^i}{h}\right)$$

High level view of density estimation

Output of training $\rightarrow f$ model

Using the model f

(Testing)

Normal $\quad \mathcal{F} = \{ f(\mid \mu, \sigma^2) \}$

Training = estimating μ^{ML}, σ^{2ML} (= formula) 2 params

Logistic $\quad \mathcal{F} = \{ f(\mid a, b), a > 0 \}$

Training = Estimating f = estimating a^{ML}, b^{ML}

gradient ascent

$O(nT)$ = T iterations $\times$ $2n$ ops

2 params

KDE $\quad \mathcal{F} = (k(\,), h)$

Training = store $k(\,), h, x^{1:n}$

n memory

effect of h

how to choose h

f fixed

x variable

⤷ new data

query points

$f(x)$ $\quad O(1)$ = constant

compute f at x

n additions

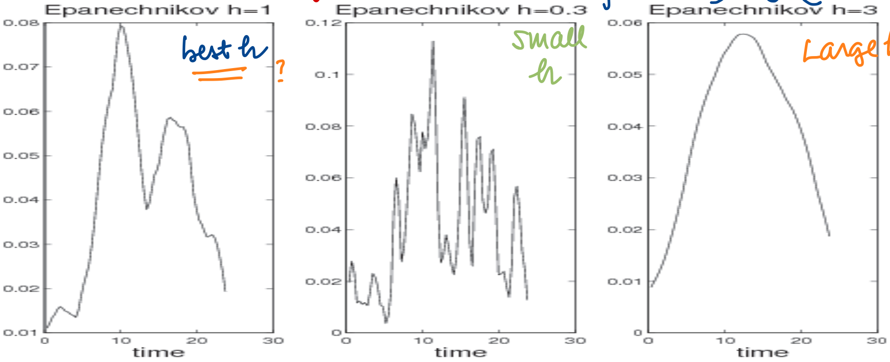
n $k(\,)$ computations

n ...

$O(n)$

~~Choosing h~~ Effect of h

$f^{true} =$ 



h = kernel width

Gaussian kernel

$$k(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$k_h(z) = \frac{1}{\sqrt{2\pi}\, h} e^{-\frac{z^2}{2h^2}} = k\left(\frac{z}{h}\right)\cdot\frac{1}{h}$$

in general

The Bias-Variance trade-off

$\boxed{\text{Choosing } h}$

NEW Principle (not ML)

Variance

$\mathcal{D} = \{x^{1:n}\} \sim iid\ f^{unknown} \rightarrow f^{model} \rightarrow$ (predictions..)

RANDOM → RANDOM → RANDOM

$Var(f(X))$

$X \sim f^{unknown}$

can't be avoided

Bias

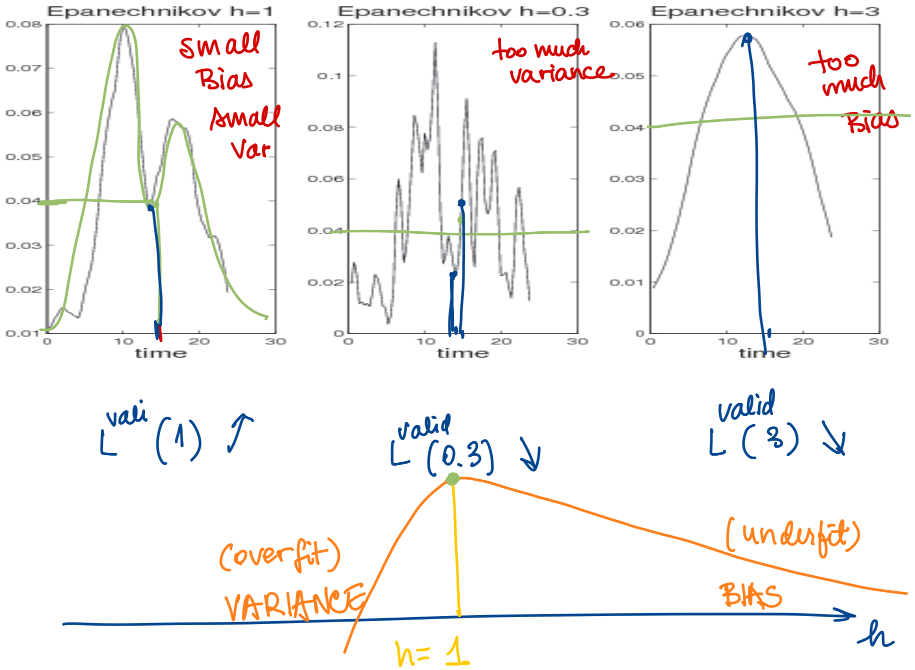
- inability of $\mathcal{M}$ to fit the data (actually f^{unkn})

can't be avoided

we must choose some $\mathcal{M}$

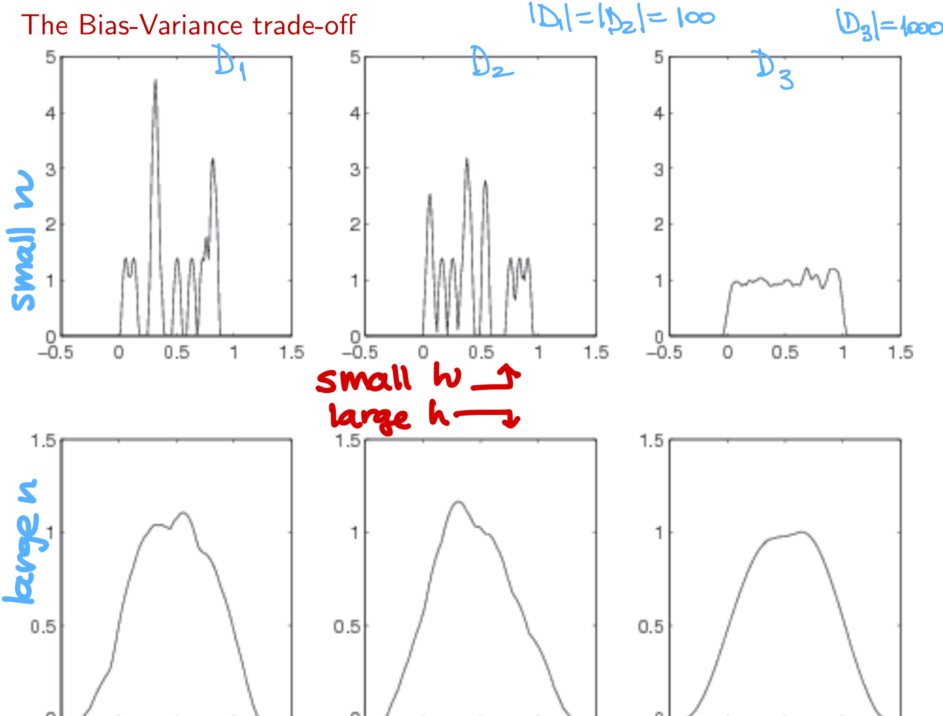
.../handouts/fig-marathon-cv2.png

Choosing h



Epanechnikov h=1
Small Bias
Small Var
time
Epanechnikov h=0.3
too much variance
time
Epanechnikov h=3
too much Bias
time
$L^{vali}(1)$
$L^{valid}(0.3)$
$L^{valid}(3)$
(overfit)
VARIANCE
(underfit)
BIAS
h
h = 1

The Bias-Variance trade-off



~~The Bias Variance trade-off~~

Choosing h in: $f_h = KDE(k(), h)$

ideal: $f_h \approx f^{unkn}$

optimize

practical: sample $\tilde{\mathcal{D}} = \tilde{x}^1, \ldots \tilde{x}^{n'} \sim f$

NEW dataset (new samples)

$$\max_h \text{Likelihood}(\underbrace{\tilde{x}^1 \ldots \tilde{x}^{n'}}_{\text{validation data}} \mid f_h)$$

training data $x^{1:n}$

CROSS-VALIDATION

for h in some set

compute $L(\tilde{\mathcal{D}} \mid f_h(\mathcal{D})) = L^{valid}(h)$

validation training

$$h^* = \arg\max_h L^{valid}(h)$$