STAT435_HW1_part2

Liyuan Tang

4/12/2020

Part a

Notice that $\hat{\mathbf{f}} = \mathbf{W}\mathbf{y} = \mathbf{W}(\mathbf{f} + \epsilon)$

D $= E(\frac{1}{n}||\mathbf{f} - \hat{\mathbf{f}}||^2)$

$= \frac{1}{n}E((\mathbf{f} - \hat{\mathbf{f}})^{\mathbf{T}}(\mathbf{f} - \hat{\mathbf{f}}))$

$= \frac{1}{n}E(\mathbf{f}^{\mathbf{T}}\mathbf{f} - \mathbf{2f}^{\mathbf{T}}\hat{\mathbf{f}} + \hat{\mathbf{f}}^{\mathbf{T}}\hat{\mathbf{f}})$

$= \frac{1}{n}E(\mathbf{f}^{\mathbf{T}}\mathbf{f} - \mathbf{2f}^{\mathbf{T}}\mathbf{W}(\mathbf{f} + \epsilon) + (\mathbf{f}^{\mathbf{T}} + \epsilon^{\mathbf{T}})\mathbf{W}^{\mathbf{T}}\mathbf{W}(\mathbf{f} + \epsilon))$

$= \frac{1}{n}E(\mathbf{f}^{\mathbf{T}}\mathbf{f} - \mathbf{2f}^{\mathbf{T}}\mathbf{W}\mathbf{f} + \mathbf{f}^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\mathbf{f}) + \frac{1}{n}E(\epsilon^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\epsilon) - \frac{1}{n}E(\mathbf{f}^{\mathbf{T}}\mathbf{W}\epsilon) + \frac{1}{n}E(\mathbf{2f}^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\epsilon)$

Consider $E(\mathbf{f}^{\mathbf{T}}\mathbf{W}\epsilon)$. Define $V = \mathbf{f}^{\mathbf{T}}\mathbf{W}$, so $E(\mathbf{f}^{\mathbf{T}}\mathbf{W}\epsilon) = E(V\epsilon) = E(\sum_i V_i\epsilon_i) = 0$ due to the linearity. And similar for the $E(\mathbf{2f}^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\epsilon)$.

So D $= \frac{1}{n}(E(\mathbf{f}^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\mathbf{f} - \mathbf{2f}^{\mathbf{T}}\mathbf{W}\mathbf{f} + \mathbf{f}^{\mathbf{T}}\mathbf{f}) + E(\epsilon^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\epsilon))$

$= \frac{1}{n}(||\mathbf{W}\mathbf{f} - \mathbf{f}||^2 + E(\epsilon^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\epsilon))$

$= \frac{1}{n}(||(\mathbf{W} - \mathbf{I})\mathbf{f}||^2 + E(\epsilon^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\epsilon))$

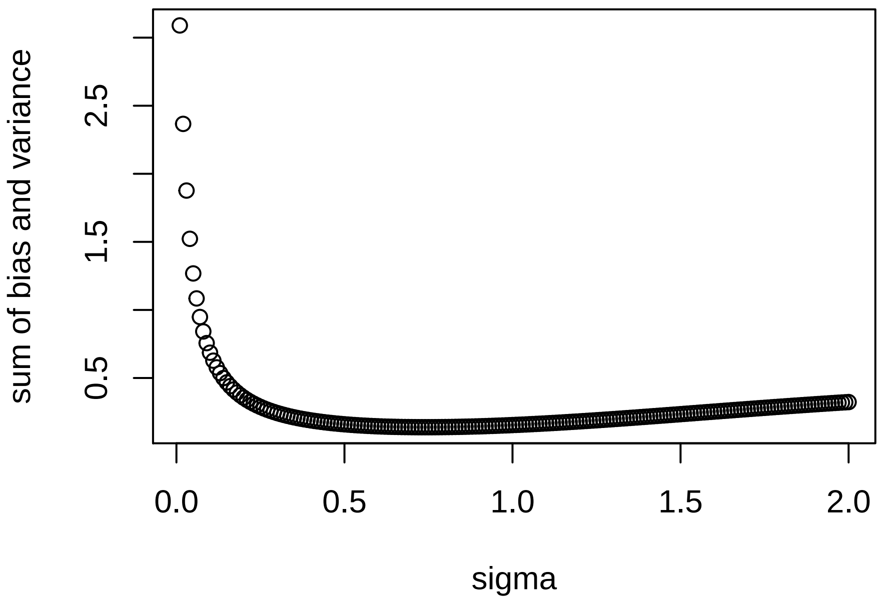
Set the vector $Y = W\epsilon$. Then $E(\epsilon^{\mathbf{T}}\mathbf{W}^{\mathbf{T}}\mathbf{W}\epsilon) = E(Y^TY) = E(\sum_{i=1}^n Y_i^2) = \sum_{i=1}^n E(Y_i^2) = \sum_{i=1}^n Var(Y_i) = \sum_{i=1}^n \sum_{j=1}^n \mathbf{W}_{ij}^2\sigma^2 = \sigma^2\mathbf{trace}(\mathbf{W}^{\mathbf{T}}\mathbf{W})$

Thus, D $= \frac{1}{n}(||(\mathbf{W} - \mathbf{I})\mathbf{f}||^2 + \sigma^2\mathbf{trace}(\mathbf{W}^{\mathbf{T}}\mathbf{W}))$

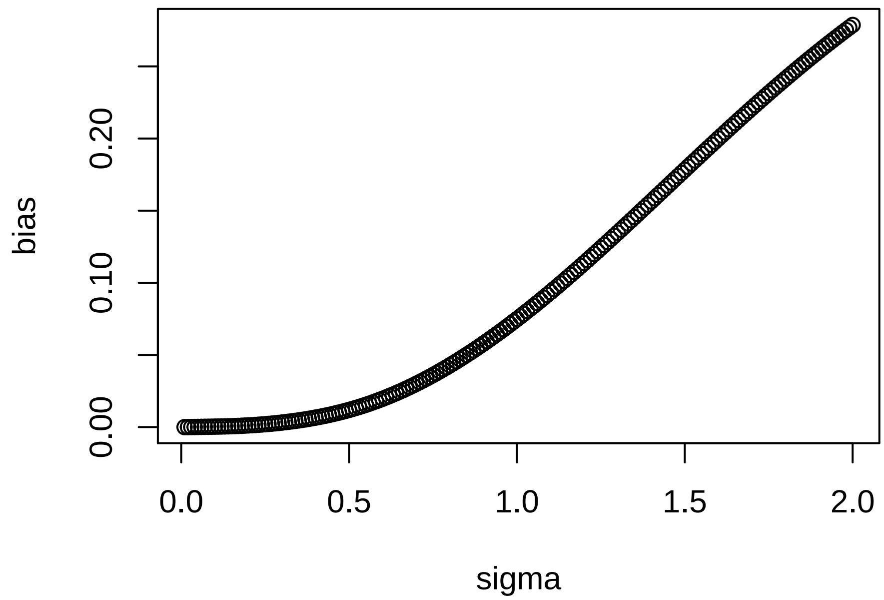
Part b

```r
source("home1-part2-data.R")
sig_vec = seq(from = 0.01, to = 2, by = 0.01)
n = length(x.train)
bias.vec = numeric(n)
var.vec = numeric(n)


for (k in 1:length(sig_vec)) {
  sig = sig_vec[k]
  # W matrix
  W.matrix = matrix(data = NA, nrow = n, ncol = n)
  for (i in 1:n) {
    bt = sum(dnorm(x.train[i] - x.train, 0 , sig))
    for (j in 1:n) {
```

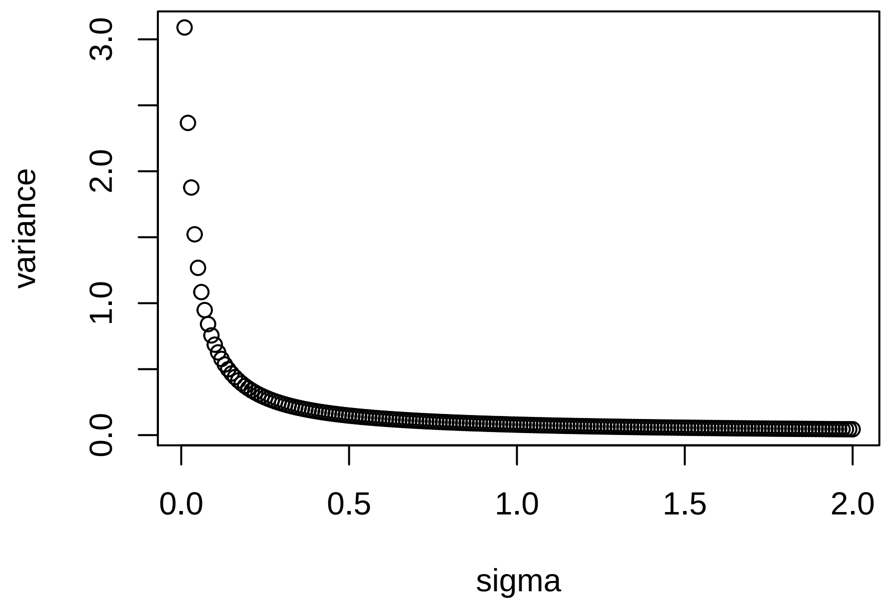
```
      W.matrix[i,j] = dnorm(x.train[i] - x.train[j], 0, sig) / bt
    }
  }

  # bias
  bias.mat = (W.matrix - diag(1,n,n)) %*% f
  bias.vec[k] = norm(x = bias.mat, type = "2")^2 / n

  # variance
  var.vec[k] = sum(diag(t(W.matrix) %*% W.matrix)) * noise.var / n
}
plot(sig_vec, bias.vec + var.vec, xlab = "sigma", ylab = "sum of bias and variance")
```




```
plot(sig_vec, bias.vec, xlab = "sigma", ylab = "bias")
```




```r
plot(sig_vec, var.vec, xlab = "sigma", ylab = "variance")
```



```
D = bias.vec + var.vec
sig_opt = sig_vec[which(D == min(D))]
cat("The optimal choice of sigma is", sig_opt)
```

```
## The optimal choice of sigma is 0.74
```

Part c

```
W.matrix = matrix(data = NA, nrow = n, ncol = n)
for (i in 1:n) {
  bt = sum(dnorm(x.train[i] - x.train, 0 , sig_opt))
  for (j in 1:n) {
    W.matrix[i,j] = dnorm(x.train[i] - x.train[j], 0, sig_opt) / bt
  }
}
fhat = W.matrix %*% y.train
plot(x.train, y.train)
lines(x.train, f,lwd = 2, col = "red")
lines(x.train, fhat, lwd = 2, col = "blue")
legend(1, -3, legend=c("f", "fhat"),
       col=c("red", "blue"), lty=c(1,1), lwd = 2)
```