Homework 3

Solutions

1. (PCA and linear regression) For PCA, we have

$$u_1 = \arg \max_{||v||_2=1} \sum_{i=1}^{n} v^T x_i. \tag{1}$$

The nearest point to x_i on the line collinear with a unit ℓ_2-norm vector v is the projection $(x_i^T v)v$, so by Pythagoras' theorem, the square of the ℓ_2-norm distance between x_i and the line equals $||x_i||_2^2 - (x_i^T v)^2$. By Eq. (1) the line collinear with u_1 minimizes the sum of square ℓ_2-norm distances.

In linear regression, the OLS coefficient is the solution to the least-squares problem

$$\min_{\beta} \sum_{i=1}^{n} (x_i[2] - \beta x_i[1])^2. \tag{2}$$

Each term is the square difference between the second component and the line. On a 2D plane, if the second component is on the vertical axis, this is the square horizontal distance between the point and the line. The line obtained from linear regression minimizes the sum of these distances. In contrast PCA minimizes the sum of square Euclidean distances as explained above, so they are different.

2. (Heartbeat)

(a) By independence

$$\mathrm{Var}(\tilde{x}[1]) = \mathrm{Var}(\tilde{b}) + \mathrm{Var}(\tilde{m}) + \mathrm{Var}(\tilde{z}_1) \tag{3}$$

$$= 12, \tag{4}$$

$$\mathrm{Cov}(\tilde{b}\tilde{x}[1]) = 1 \tag{5}$$

so

$$\Sigma_{\tilde{x}[1]} = 12 \tag{6}$$

$$\Sigma_{\tilde{b}\tilde{x}[1]} = 1 \tag{7}$$

and by Theorem 2.3 in the notes the estimate equals

$$\hat{b}(\tilde{x}) = \frac{\tilde{x}[1]}{12}, \tag{8}$$

and the corresponding MSE equals

$$\mathrm{E}((\hat{b}(\tilde{x}) - \tilde{b})^2) = \mathrm{Var}(\tilde{b}) - \frac{1}{12} \tag{9}$$

$$= 0.92. \tag{10}$$

The estimate just shrinks the signal.

(b) By independence

$$\begin{aligned}
\mathrm{Var}(\tilde{x}[1]) &= \mathrm{Var}(\tilde{b}) + \mathrm{Var}(\tilde{m}) + \mathrm{Var}(\tilde{z}_1) && (11)\\
&= 12, && (12)\\
\mathrm{Var}(\tilde{x}[2]) &= \mathrm{Var}(\tilde{m}) + \mathrm{Var}(\tilde{z}_2) && (13)\\
&= 11, && (14)\\
\mathrm{Cov}(\tilde{x}[1]\tilde{x}[2]) &= \mathrm{E}(\tilde{m}^2) && (15)\\
&= 10, && (16)\\
\mathrm{Cov}(\tilde{b}\tilde{x}[1]) &= 1 && (17)\\
\mathrm{Cov}(\tilde{b}\tilde{x}[2]) &= 0 && (18)\\
& && (19)
\end{aligned}$$

so

$$\begin{aligned}
\Sigma_{\tilde{x}} &= \begin{bmatrix} 12 & 10 \\ 10 & 11 \end{bmatrix} && (20)\\
\Sigma_{\tilde{b}\tilde{x}} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} && (21)
\end{aligned}$$

and by Theorem 2.3 in the notes the estimate equals

$$\begin{aligned}
\hat{b}(\tilde{x}) &= \tilde{x}^T \begin{bmatrix} 12 & 10 \\ 10 & 11 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} && (22)\\
&= \frac{11\tilde{x}[1] - 10\tilde{x}[2]}{32}, && (23)
\end{aligned}$$

and the corresponding MSE equals

$$\begin{aligned}
\mathrm{E}((\hat{b}(\tilde{x}) - \tilde{b})^2) &= \mathrm{Var}(\tilde{b}) - \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 12 & 10 \\ 10 & 11 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} && (24)\\
&= 1 - \frac{11}{32} && (25)\\
&= 0.66. && (26)
\end{aligned}$$

The estimate approximately cancels out the signal from the mum, in a way that is optimal with respect to MSE.

3. (Gaussian minimum MSE estimator)

(a) As explained in Section 2 in the PCA notes, the covariance is a valid inner product for zero-mean random variables, so *rndc* is the component of $\tilde{b}$ collinear with $\tilde{a}$ and $\tilde{b} - \tilde{c}$ is the component of $\tilde{b}$ orthogonal to $\tilde{a}$.

(b) $\tilde{c}$ is a function of $\tilde{a}$, so the conditional expectation just equals $\frac{\mathrm{Cov}(\tilde{a},\tilde{b})}{\mathrm{Var}(\tilde{a})} a$.

(c) As explained in the first answer, $\tilde{b} - \tilde{c}$ is orthogonal to $\tilde{a}$ if we interpret covariance as an inner product. Indeed,

$$\begin{aligned}
\mathrm{E}(\tilde{a}(\tilde{b} - \tilde{c})) &= \mathrm{E}(\tilde{a}\tilde{b}) - \mathrm{E}(\tilde{a}\tilde{c}) && (27)\\
&= \mathrm{E}(\tilde{a}\tilde{b}) - \frac{\mathrm{Cov}(\tilde{a}, \tilde{b})}{\mathrm{Var}(\tilde{a})}\mathrm{Var}(\tilde{a}) && (28)\\
&= 0. && (29)
\end{aligned}$$

Because the random variables are Gaussian, this implies that they are independent. As a result the conditional expectation of $\tilde{b} - \tilde{c}$ given $\tilde{a} = a$ is just $\mathrm{E}(\tilde{b} - \tilde{c}) = 0$.

(d) The conditional expectation of $\tilde{b}$ given $\tilde{a} = a$ for any fixed $a \in \mathbb{R}$ equals

$$\begin{aligned}
\mathrm{E}(\tilde{b} \,|\, \tilde{a} = a) &= \mathrm{E}(\tilde{c} \,|\, \tilde{a} = a) + \mathrm{E}(\tilde{b} - \tilde{c} \,|\, \tilde{a} = a) && (30)\\
&= \frac{\mathrm{Cov}(\tilde{a}, \tilde{b})}{\mathrm{Var}(\tilde{a})} a. && (31)
\end{aligned}$$

Since the conditional mean is the minimum MSE estimate, this proves the result.

(e) Uncorrelation does not imply independence for non-Gaussian random variables.

4. (Oxford Dataset)

   (a) Figure 1

   (b) Figure 2

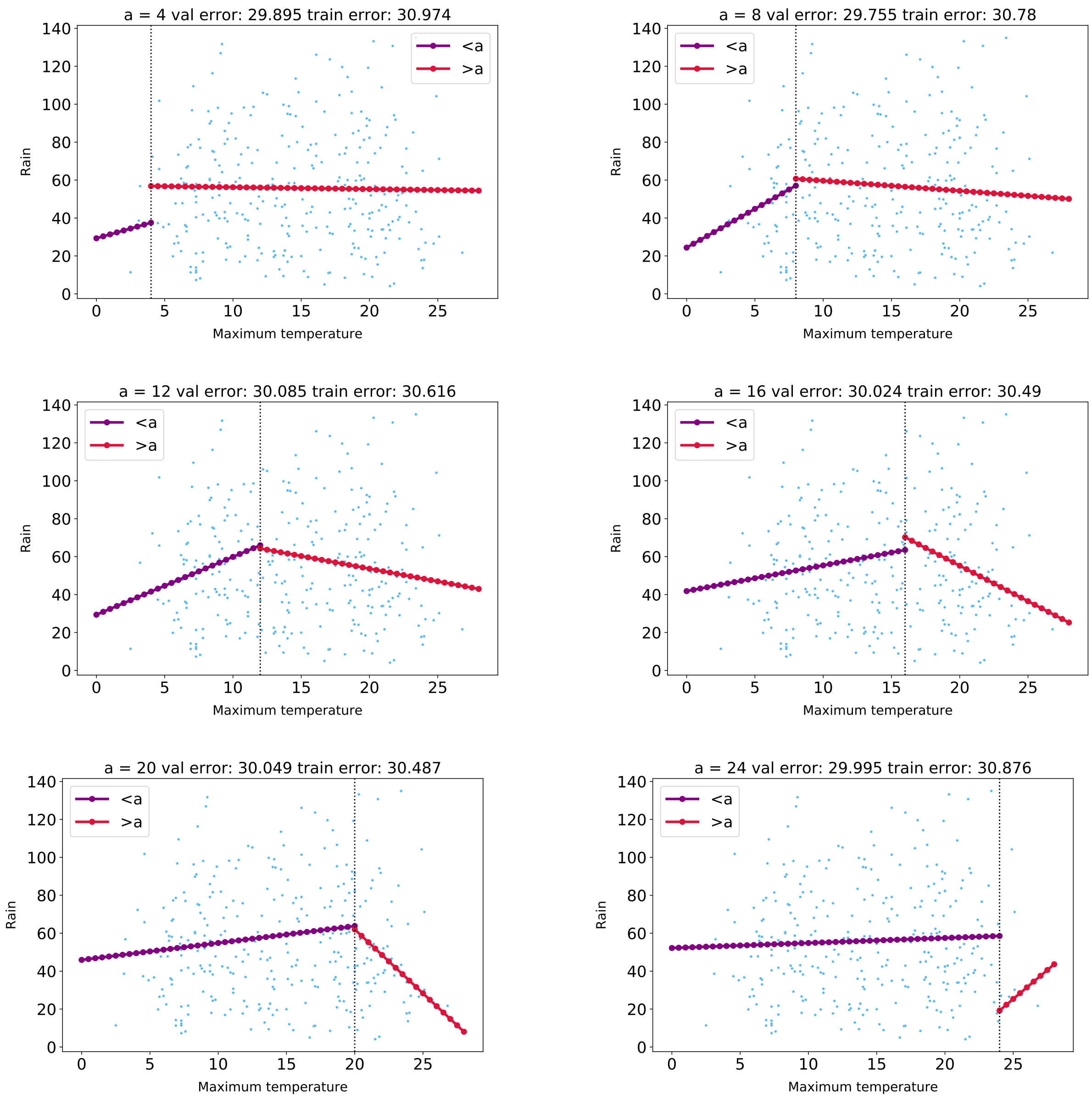

Figure 1: Fit for different values of a

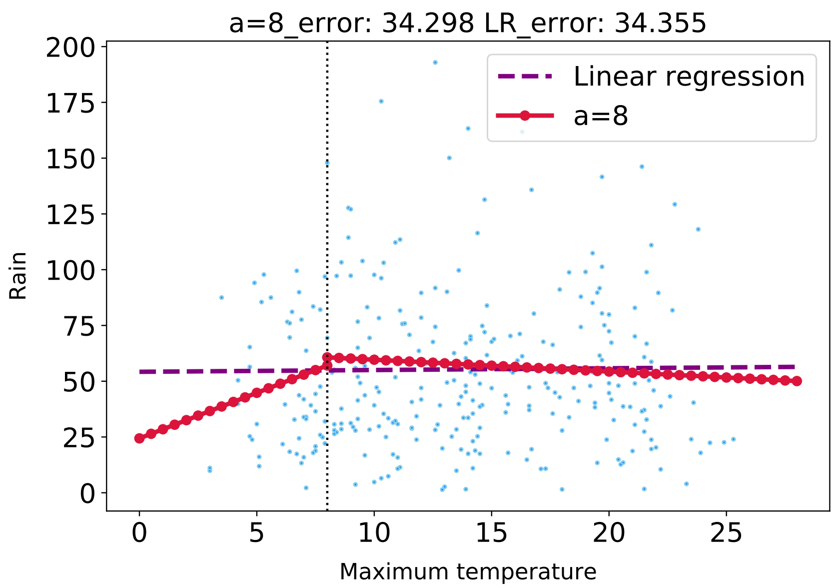

Figure 2: Comparison of best estimator in 4(a) with linear regression on the test dataset