4/20/20

Regression Quality Metrics

trained model: $model(\tilde{x}, \tilde{w}^*) = \tilde{x}^T \tilde{w}^*$

↑ ↑
LS LAD

prediction:

$\tilde{x}^0 : model(\tilde{x}^0, \tilde{w}^*) = y_0$

Judging the quality of trained model

→ $\underline{MSE} = \frac{1}{P} \sum_{p=1}^{P} \left(model(\tilde{x}_p, \tilde{w}^*) - y_p \right)^2$

↑
LS, LAD

→ $MAD = \frac{1}{P} \sum_{p=1}^{P} \left| model(\tilde{x}_p, \tilde{w}^*) - y_p \right|$

min absolute deviations

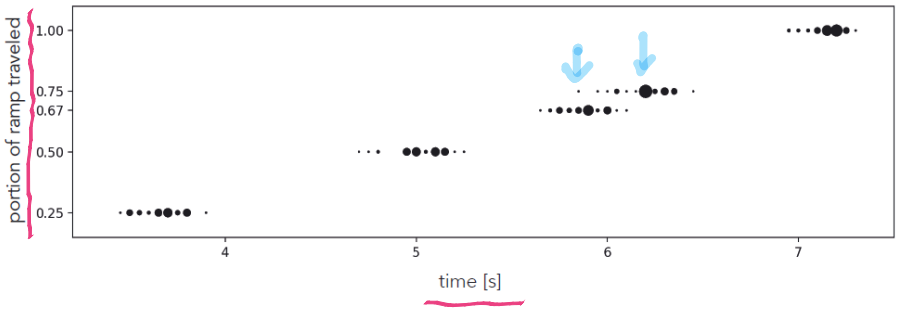


Figure 5.9 Figure associated with Example 5.7. See text for details.

Weighted Regression

emphasize or de-emphasize points

- deal w/ duplicates

due to quantization or binning

assume b_p is the number of identical in/out pair $(\tilde{x}_p, y_p)$

Error

$$\underbrace{\left(\text{model}(\tilde{x}_1, \tilde{w}) - y_1\right)^2 + \cdots + \left(\text{model}(\tilde{x}_1, \tilde{w}) - y_1\right)^2}_{b_1}$$

$$+ \quad \cdots \quad + \underbrace{\left(\text{model}(\tilde{x}_p, \tilde{w}) - y_p\right)^2 + \cdots + \left(\text{model}(\tilde{x}_p, \tilde{w}) - y_p\right)^2}_{b_p}$$

$$\therefore \quad g(\tilde{w}) = \frac{1}{b_1 + b_2 + \cdots + b_p} \sum_{p=1}^{P} b_p \left(\text{model}(\tilde{x}_p, \tilde{w}) - y_p\right)^2$$

- Weight points by confidence

  confidence in the trustworthiness of each data point

Fig 5.10

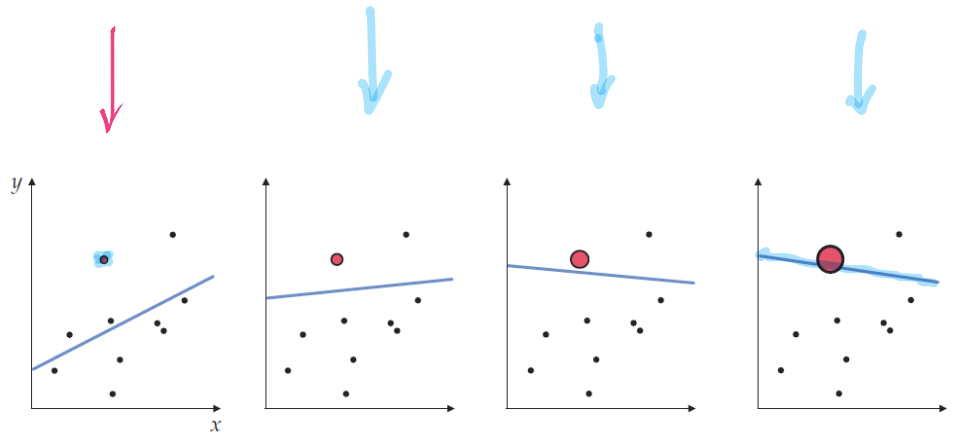


Figure 5.10 Figure associated with Example 5.8. See text for details.

Multi-output Regression

both input & output are vector-valued

P input pairs

$$(\bar{x}_1, \bar{y}_1), \ldots, (\bar{x}_P, \bar{y}_P) \qquad \bar{x}_i \in R^N \quad \bar{y}_i \in R^C$$

$$\bar{x}_p = \begin{bmatrix} x_{1,p} \\ x_{2,p} \\ \vdots \\ x_{N,p} \end{bmatrix}_{N \times 1}, \quad \bar{y}_p = [y_{0,p} \; y_{1,p} \cdots y_{c,p} \cdots y_{C-1,p}]_{1 \times C}$$

assume: a linear relationship holds between $\bar{x}_p$ & just the c-th dimension of $\bar{y}_p$

$$\longrightarrow \boxed{\tilde{x}_p^T \tilde{w}_c \approx y_{c,p}}, \quad p = 1, \ldots, P$$

scalar

$$\tilde{w}_c = [b_c \; w_{1,c} \cdots w_{N,c}]^T$$

assume: above applies to all C entries of output

$$W = \begin{bmatrix} \tilde{w}_0 & \tilde{w}_1 & \tilde{w}_c & \tilde{w}_{C-1} \\ b_0 & b_1 & b_c & b_{C-1} \\ w_{1,0} & w_{1,1} & w_{1,c} & w_{1,C-1} \\ \vdots & \vdots & \vdots & \vdots \\ w_{N,0} & w_{N,1} & w_{N,c} & w_{N,C-1} \end{bmatrix} \quad (N+1) \times C$$

$$\tilde{x}_p^T W = \begin{bmatrix} \tilde{x}_p^T \tilde{w}_0 & \tilde{x}_p^T \tilde{w}_1 & \cdots & \tilde{x}_p^T \tilde{w}_{C-1} \end{bmatrix}$$

$$\boxed{\tilde{x}_p^T W \simeq \bar{y}_p \,, \quad p = 1, \ldots, P}$$

$1 \times (N+1) \quad (N+1) \times C \quad 1 \times C$

error at the p-th point : $\underline{(\tilde{x}_p^T W - \bar{y}_p)}$ $1 \times C$

ℓ_2 : Least squares:

$$g(W) = \frac{1}{P} \sum_{p=1}^{P} \| \tilde{x}_p^T W - \bar{y}_p \|_2^2 = \frac{1}{P} \sum_{p=1}^{P} \sum_{c=0}^{C-1} (\tilde{x}_p^T \tilde{w}_c - y_{c,p})^2$$

ℓ_1 : LAD

$$g(W) = \frac{1}{P} \sum_{p=1}^{P} \| \tilde{x}_p^T W - \bar{y}_p \|_1 = \frac{1}{P} \sum_{p=1}^{P} \sum_{c=0}^{C-1} | \tilde{x}_p^T \tilde{w}_c - y_{c,p} |$$

$$\underline{\underline{g(W)}} = \sum_{c=0}^{C-1} \underbrace{\left(\frac{1}{P} \sum_{p=1}^{P} | \tilde{x}_p^T \tilde{w}_c - y_{c,p} | \right)}_{g_c(\tilde{w}_c)} = \sum_{c=0}^{C-1} g_c(\tilde{w}_c)$$

$g_c(\underline{\tilde{w}_c})$

Independent optimization for each $\tilde{w}_c$

$N = 2, \quad C = 2$

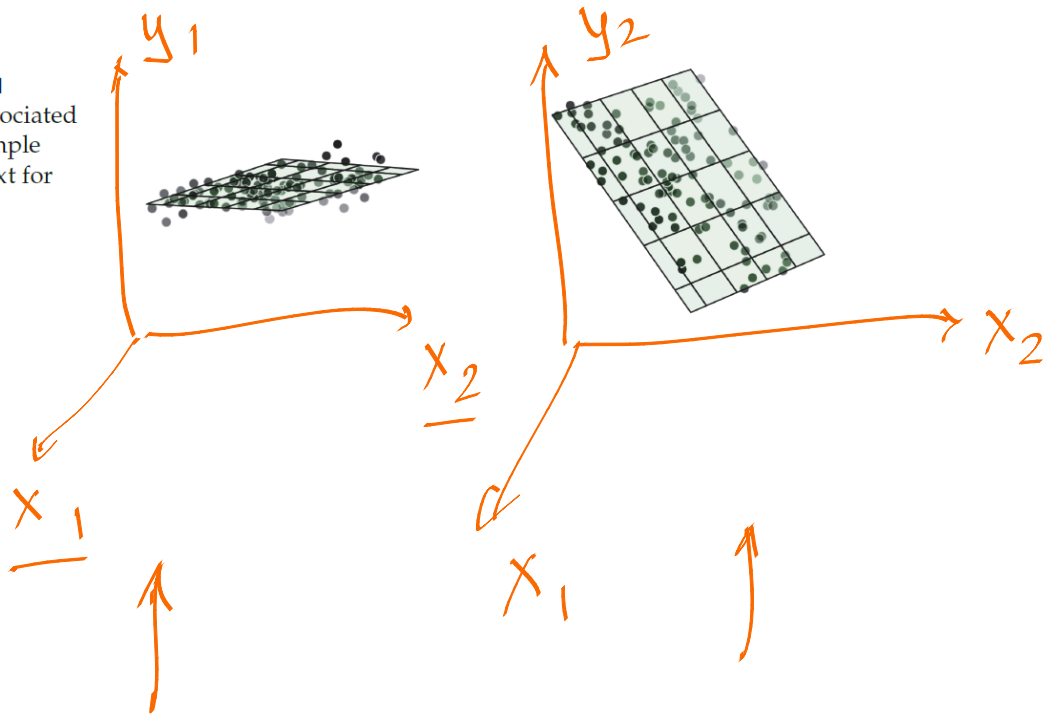


Figure 5.11
Figure associated with Example 5.9. See text for details.

Vector Norms

$\bar{x} \in R^N \qquad \bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$

Euclidean norm or ℓ_2 norm

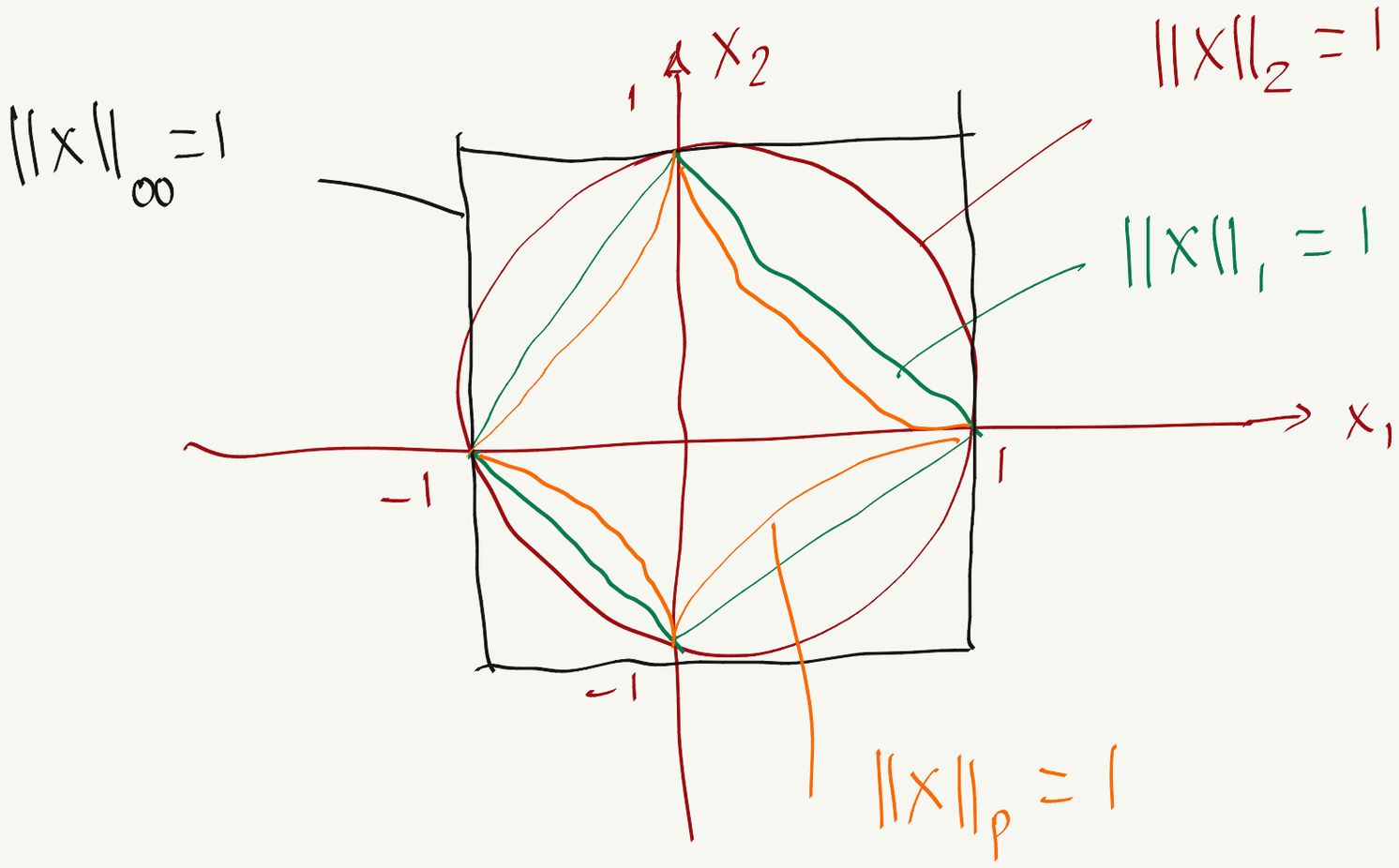
$\ell_2: \quad \|x\|_2 = \sqrt{\sum_{n=1}^{N} x_n^2}$

$\ell_1: \quad \|x\|_1 = \sum_{n=1}^{N} |x_n|$

$\ell_\infty: \quad \|x\|_\infty = \max_n |x_n|$

$\ell_p: \quad \|x\|_p = \left(\sum_{n=1}^{N} |x_n|^p \right)^{1/p}$

UNIT BALL



In problem 5.9 in the homework you are asked to do standard normalization of the input-features

Standard Normalization (9.3)

how? why?

$\bar{x}_1$ $\bar{x}_2$ $\bar{x}_p$ $\bar{x}_P$

$$\begin{matrix} x_{1,1} \\ \vdots \\ x_{1,n} \\ \vdots \\ x_{1,N} \end{matrix} \quad \begin{matrix} x_{2,1} \\ \\ x_{2,n} \\ \\ x_{2,N} \end{matrix} \quad \cdots \quad \begin{matrix} x_{p,1} \\ \\ x_{p,n} \\ \\ x_{p,N} \end{matrix} \quad \cdots \quad \begin{matrix} x_{P,1} \\ \vdots \\ x_{P,n} \\ \vdots \\ x_{P,N} \end{matrix}$$

$\mu_n, \sigma_n \rightarrow$ n-th input feature

invertible

$$x_{p,n} \leftarrow \frac{x_{p,n} - \mu_n}{\sigma_n}, \quad n = 1, \dots, N$$

$$\mu_n = \frac{1}{P} \sum_{p=1}^{P} x_{p,n}$$

$$\sigma_n = \left(\frac{1}{P} \sum_{p=1}^{P} (x_{p,n} - \mu_n)^2 \right)^{1/2}$$

except when $\sigma_n = 0$ (this means $x_{i,n} = \text{constant}$)

the n-th feature is redundant

$\Rightarrow$ is removed

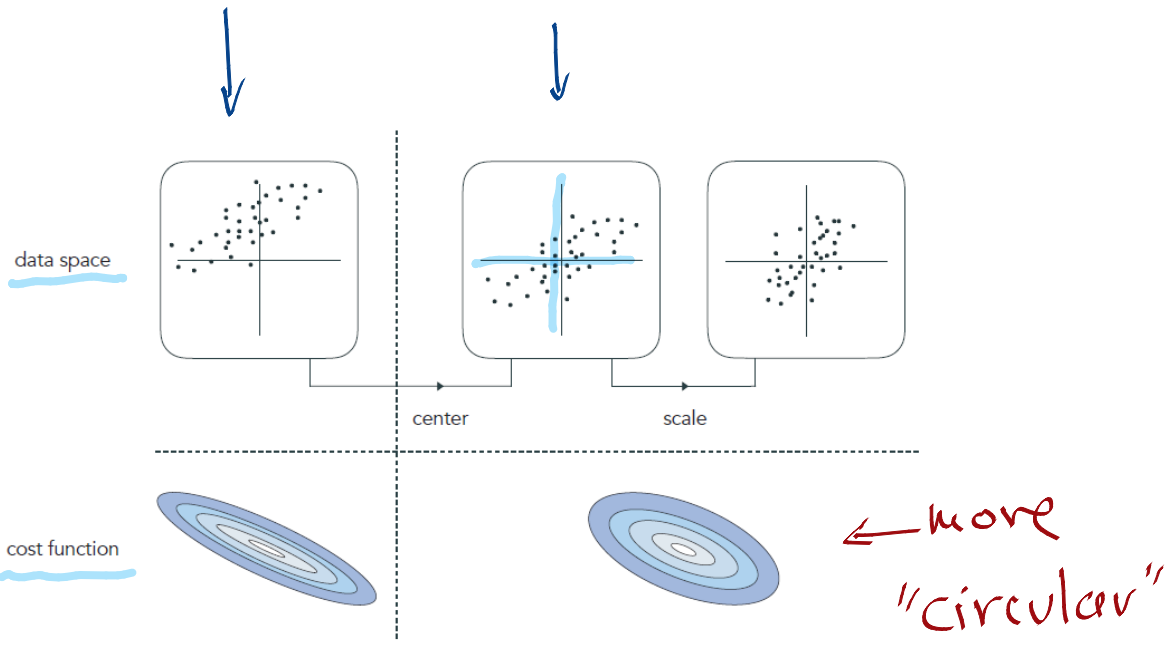


Figure 9.14 (Standard normalization illustrated. The input space of a generic dataset (top-left panel) as well as its mean-centered (top-middle panel) and scaled version (top-right panel). As illustrated in the bottom row, where a prototypical cost function corresponding to this data is shown, standard normalization results in a cost function with less elliptical and more circular contours compared to the original cost function.

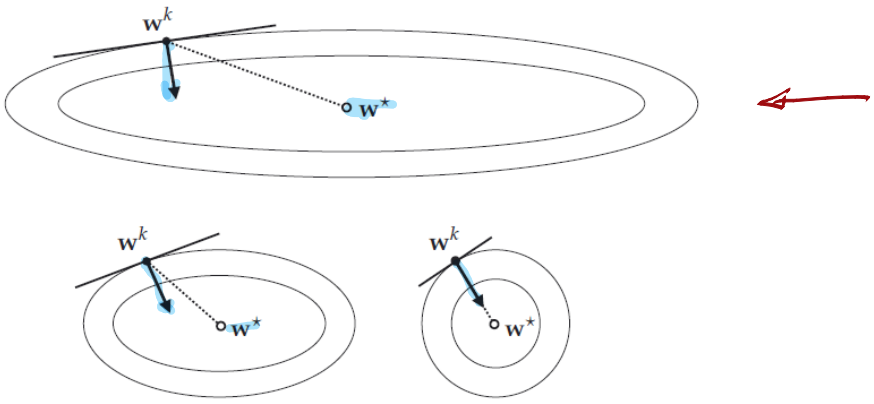
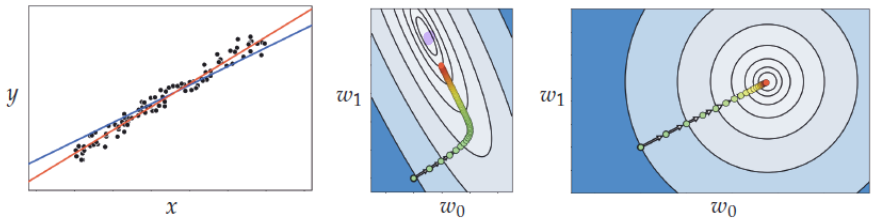


Figure 9.15 In standard normalizing input data we temper its associated cost function's often-elliptical contours, like those shown in the top panel, into more circular ones as shown in the bottom-left and bottom-right panels. This means that the gradient descent direction, which points away from the minimizer of a cost function when its contours are elliptical (leading to the common zig-zagging problem with gradient descent), points more towards the function's minimizer as its contours become more circular. This makes each gradient descent step much more effective, typically allowing the use of much larger steplength parameter values α, meaning that measurably fewer steps are required to adequately minimize the cost function.

Speed up convergence
(when GD is used)

No normalization

with normalization



100 it

20 it

$\bar{w}^0 = [0 \quad 0]^T$

$\alpha = 10^{-1}$

blue line

red line