Solution 1:

a) multiclass classification (plate digits) (supervised learning)

b) binary classification (supervised)

c) outlier detection ((un)supervised)

d) frequent pattern mining (unsupervised)

e) classification (supervised) / clustering (unsupervised)

f) classification (supervised)

g) clustering / assocation rules (unsupervised)

h) not a machine learning task

i) not a machine learning task

Solution 2:

a) We use the least squares-estimator introduced in the lecture: $\hat{\beta} = (X^T X)^{-1} X^T y$ with

$$X = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & ... & x_{1,m} \\ 2 & x_{2,1} & x_{2,2} & ... & x_{2,m} \\ \vdots & \vdots & \vdots & ... & \vdots \\ n & x_{n,1} & x_{n,2} & ... & x_{n,m} \end{bmatrix}$$

$$x = \begin{bmatrix} 0.56 \\ 0.22 \\ 1.7 \\ 0.63 \\ 0.36 \\ 1.2 \end{bmatrix}, X = \begin{bmatrix} 1 & 0.56 \\ 1 & 0.22 \\ 1 & 1.7 \\ 1 & 0.63 \\ 1 & 0.36 \\ 1 & 1.2 \end{bmatrix} \text{ and } y = \begin{bmatrix} 160 \\ 150 \\ 175 \\ 185 \\ 165 \\ 170 \end{bmatrix}$$

Then

$\hat{\beta} = (X^T X)^{-1} X^T y$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & ... & 1 \\ x_{1,1} & x_{2,1} & x_{3,1} & ... & x_{n,1} \\ \vdots & \vdots & \vdots & ... & \vdots \\ x_{1,m} & x_{2,m} & x_{3,m} & ... & x_{n,m} \end{bmatrix} \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & ... & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & ... & x_{2,m} \\ \vdots & \vdots & \vdots & ... & \vdots \\ 1 & x_{n,1} & x_{n,2} & ... & x_{n,m} \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & ...1 & \\ x_{1,1} & x_{2,1} & x_{3,1} & ... & x_{n,1} \\ \vdots & \vdots & \vdots & ... & \vdots \\ x_{1,m} & x_{2,m} & x_{3,m} & ... & x_{n,m} \end{bmatrix} \begin{bmatrix} 160 \\ 150 \\ 175 \\ 185 \\ 165 \\ 170 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.56 & 0.22 & 1.7 & 0.63 & 0.36 & 1.2 \end{bmatrix} \begin{bmatrix} 1 & 0.56 \\ 1 & 0.22 \\ 1 & 1.7 \\ 1 & 0.63 \\ 1 & 0.36 \\ 1 & 1.2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.56 & 0.22 & 1.7 & 0.63 & 0.36 & 1.2 \end{bmatrix} \begin{bmatrix} 160 \\ 150 \\ 175 \\ 185 \\ 165 \\ 170 \end{bmatrix}$$

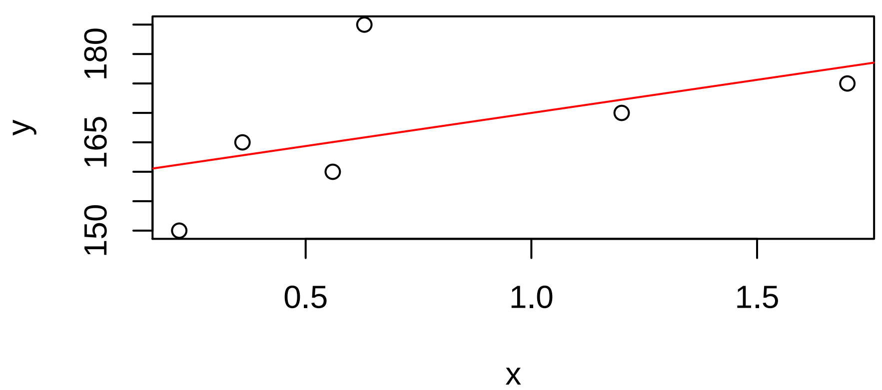
$$= \begin{bmatrix} 6 & 4.67 \\ 4.67 & 5.2185 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.56 & 0.22 & 1.7 & 0.63 & 0.36 & 1.2 \end{bmatrix} \begin{bmatrix} 160 \\ 150 \\ 175 \\ 185 \\ 165 \\ 170 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5491944 & -0.4914703 \\ -0.4914703 & 0.6314394 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.56 & 0.22 & 1.7 & 0.63 & 0.36 & 1.2 \end{bmatrix} \begin{bmatrix} 160 \\ 150 \\ 175 \\ 185 \\ 165 \\ 170 \end{bmatrix}$$

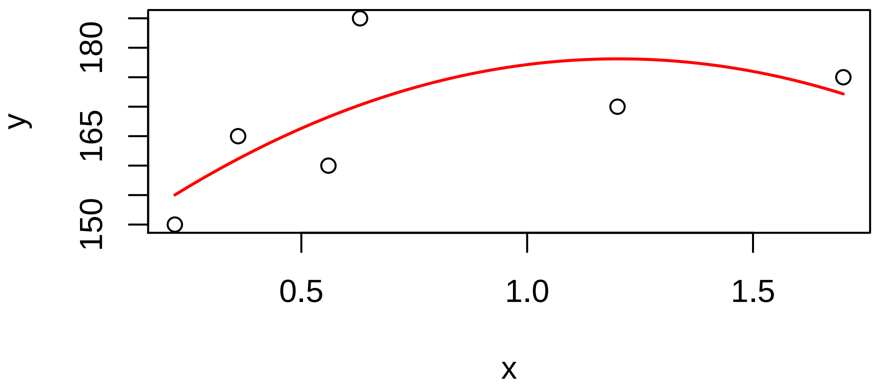
$$= \begin{bmatrix} 0.2739710 & 0.4410709 & -0.2863051 & 0.23956809 & 0.3722651 & -0.04056998 \\ -0.1378643 & -0.3525536 & 0.5819766 & -0.09366351 & -0.2641521 & 0.26625693 \end{bmatrix} \begin{bmatrix} 160 \\ 150 \\ 175 \\ 185 \\ 165 \\ 170 \end{bmatrix}$$

$$= \begin{bmatrix} 158.73954 \\ 11.25541 \end{bmatrix}$$

Hence the linear model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 158.73954 + 11.25541x$



b) Here $X = \begin{bmatrix} 1 & 0.56 & 0.3136 \\ 1 & 0.22 & 0.0484 \\ 1 & 1.7 & 2.89 \\ 1 & 0.63 & 0.3969 \\ 1 & 0.36 & 0.1296 \\ 1 & 1.2 & 1.44 \end{bmatrix}$ and $\hat{\beta} = \begin{bmatrix} 143.51682 \\ 57.59155 \\ -23.96347 \end{bmatrix}$



Solution 3:

$$
\begin{aligned}
L(c) &= \sum_{i=1}^{n} |y^{(i)} - c| \\
&= \sum_{i=1}^{n} (y^{(i)} - c)\mathbb{1}_{\{y^{(i)} > c\}} + \sum_{i=1}^{n} (c - y^{(i)})\mathbb{1}_{\{y^{(i)} \le c\}} \\
&= \sum_{i=1}^{n} y^{(i)} \left(\mathbb{1}_{\{y^{(i)} > c\}} - \mathbb{1}_{\{y^{(i)} \le c\}}\right) + c \sum_{i=1}^{n} \left(\mathbb{1}_{\{y^{(i)} \le c\}} - \mathbb{1}_{\{y^{(i)} > c\}}\right)
\end{aligned}
$$

$L(c)$ is piecewise linear, so $L'(c)$ is piecewise continuous with jumps at $\{y^{(1)}, ..., y^{(n)}\}$. So for $c \in \mathbb{R}_{\setminus \{y^{(1)}, .., y^{(n)}\}}$

$$
\begin{aligned}
L'(c) &= \sum_{i=1}^{n} \left(\mathbb{1}_{\{y^{(i)} \le c\}} - \mathbb{1}_{\{y^{(i)} > c\}}\right) \\
&= \left\{ \begin{array}{ll} < 0, & \text{when more than half of the y's are bigger then c} \\ > 0, & \text{when more than half of the y's are smaller then c} \end{array} \right\} \\
&= \left\{ \begin{array}{ll} < 0, & \text{for } c \le y^{(\frac{n}{2})} \text{for n even or c} < y^{(\frac{n+1}{2})} \text{for n odd} \\ > 0, & \text{for } c \ge y^{(\frac{n}{2}+1)} \text{for n even or c} > y^{(\frac{n+1}{2})} \text{for n odd} \end{array} \right\}
\end{aligned}
$$

Minimum for $L(c)$ is for the point at which $L'(c) = 0$. So we can conclude, that

for even n

$$
\begin{cases} L'(c) < 0, & \text{for } c \le y^{(\frac{n}{2})} \\ L'(c) > 0, & \text{for } c \ge y^{(\frac{n}{2}+1)} \end{cases}
$$

so c $\in \left(y^{(\frac{n}{2})}, y^{(\frac{n}{2}+1)}\right)$.

for odd n

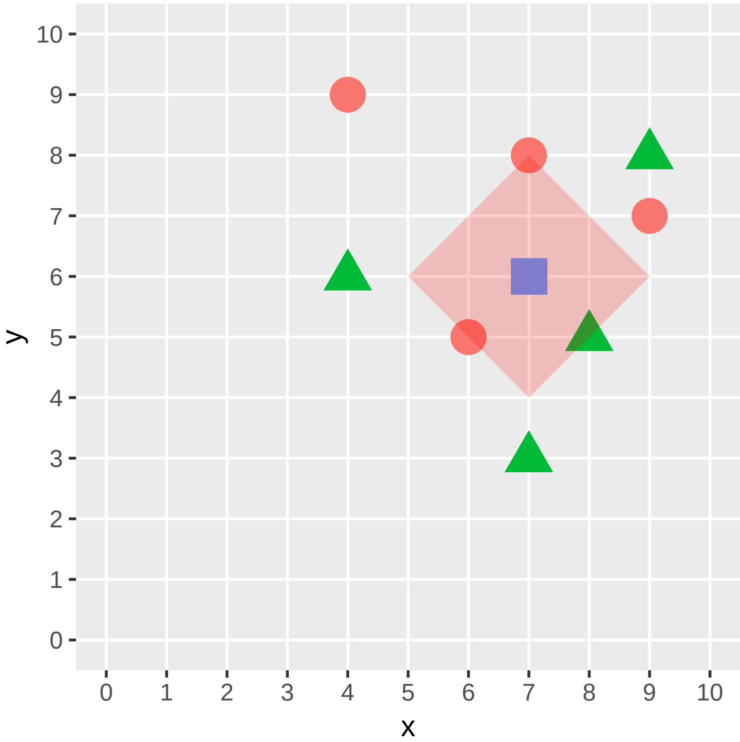
$$\begin{cases} L'(c) < 0, \text{ for } c < y^{(\frac{n+1}{2})} \\ L'(c) > 0, \text{ for } c > y^{(\frac{n+1}{2})} \end{cases}$$

so c $= y^{(\frac{n+1}{2})}$.

Solution 4:

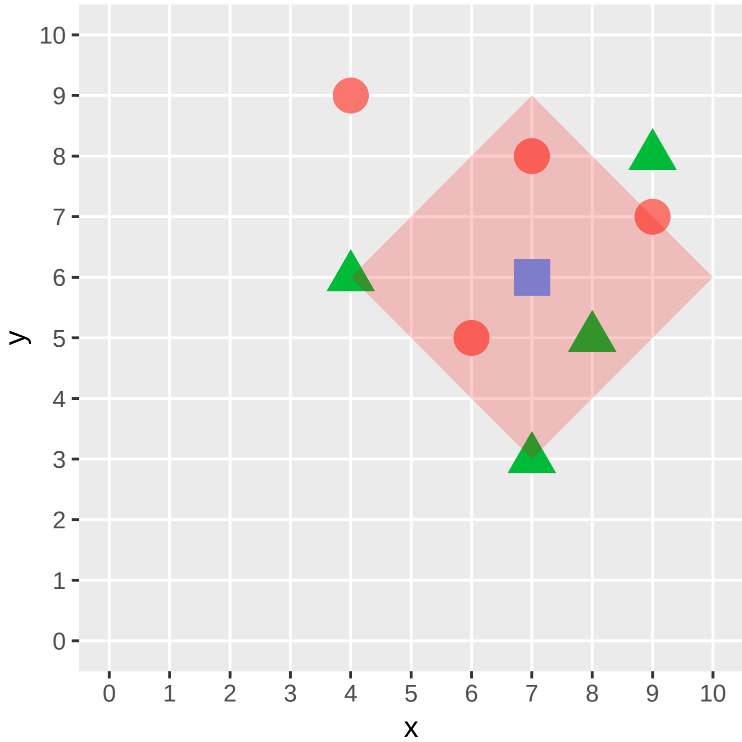
a) $k = 3$

2 circles and 1 triangle, so our point is also a circle



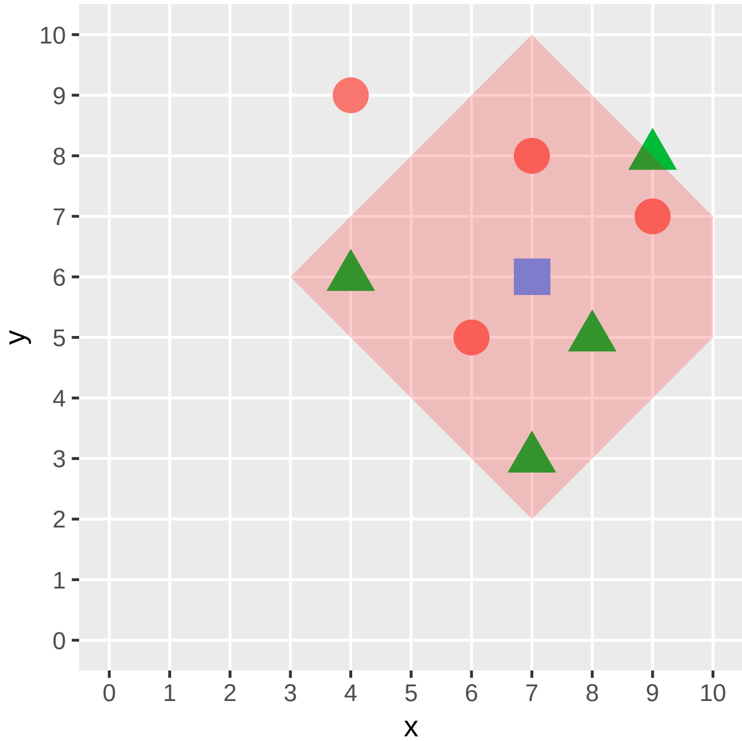
b) $k = 5$

3 circles and 3 triangles, we have to specify beforehand what to do in case of a tie



c) $k = 7$

3 circles and 4 triangles, so our point is also a triangle



Solution 5:
See R code **lm_knn_l_1a.R**