Solution 1:

a) Derivation of Gower and feature-wise distances:

$\text{Sex}^{(1)}$	$\text{Sex}^{(2)}$	$\delta_G(\text{Sex}^{(1)}, \text{Sex}^{(2)})$	$\text{Age}^{(1)}$	$\text{Age}^{(2)}$	$\delta_G(\text{Age}^{(1)}, \text{Age}^{(2)})$	$d_G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$
F	F	0	15	15	0	0
F	F	0	15	58	$= \frac{1}{90-15}\|15 - 58\| = 0.573$	0.287
F	F	0	15	90	1	0.5
F	M	1	15	15	0	0.5
F	M	1	15	58	0.573	0.787
F	M	1	15	90	1	1

b) The table shows that the Gower distance tends to favor the observations sharing the same sex. Two units having the same sex but largely differ in their age are considered closer than two units having the same age but a different gender. A feature-wise distance of 1 is only achieved for age if it has extreme values (90), which is rarely the case, since many other values could be achieved in between. While the only possible flip in sex from female to male, always leads to the maximum feature-wise distance of 1 - independent of the number of classes of the categorical features. Outliers in age could further aggravate this issue.

c) Possible weighting schemes:

- User-defined weighting schemes depending on the use case: features that should not change receive high weights.
- Weight categorical features by their number of categories, such that changes in categorical features with many classes lead to lower distances.
- Use the weighting scheme of exercise sheet 5, exercise 2 c) to find a ranking of the nominal scaled features using multi-dimensional scaling. Rank them and treat them as numerical features, where R_j is equal to the maximum rank.
- ...

Solution 2:
LIME: Example

a) Filled out table:

TO DO: Are these correct? Students reported 0.60 and 0.52, TO CHECK

	pension	age	job type	marital status	$\hat{f}$	$d(\mathbf{x}, \mathbf{z}.)$	$\phi_{\sigma=0.15}(\mathbf{z}.)$	$\phi_{\sigma=0.5}(\mathbf{z}.)$
$\mathbf{x}$	1800	21	sedentary	single	30.6	-	-	-
$\mathbf{z}_1$	1600	21	sedentary	married	25.8	0.25	0.06	0.78
$\mathbf{z}_3$	2200	32	sedentary	married	85.2	0.32	0.01	0.66
$\mathbf{z}_2$	1200	23	physically	single	74.9	0.49	0.00	0.38

- The smaller the kernel width σ the smaller the proximity measure, the smaller the weight for the sampled data points
- If the kernel is set too small, many or all sampled observations receive a weight close to 0.
- Since there are not many data points used to fit the surrogate model, the model might be unstable and not faithful to the original model.

b)

$$
\begin{aligned}
L(\hat{f}, g_1, \phi_{\mathbf{x}}) &= \sum_{\mathbf{z} \in Z} \phi_{\mathbf{x}}(\mathbf{z}) L(\hat{f}(\mathbf{z}), g(\mathbf{z})) \\
&= 0.06 \cdot (28 - 25.8)^2 + 0.01 \cdot (105 - 85.2)^2 + 0 \\
&= 4.21
\end{aligned}
$$

$$
\begin{aligned}
L(\hat{f}, g_2, \phi_{\mathbf{x}}) &= \sum_{\mathbf{z} \in Z} \phi_{\mathbf{x}}(\mathbf{z}) L(\hat{f}(\mathbf{z}), g(\mathbf{z})) \\
&= 0.06 \cdot (26.1 - 25.8)^2 + 0.01 \cdot (92.7 - 85.2)^2 + 0 \\
&= 0.57
\end{aligned}
$$

According to the faithfulness, g_2 should be preferred because it has a lower weighted loss.

c) No, because a random forest is by far less interpretable than a linear model with three features.

d) Yes, because there is a high probability that the random forest overfitted on the sampled data. With a new sampled dataset the faithfulness might be lower for the random forest.

As discussed in the exercise session, exercise 3 of this week on LIME implementation is still left open and will be discussed in next week's exercise class.