sheldon

smart habitat
for the elderly

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the H2020 Framework Programme
of the European Union

This publication is based upon work from COST Action CA16226: Indoor Living Space Improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.**cost**.eu

www.**sheld-on**.eu

Course:

Explainable AI in AAL

Lecture 2:

SHAP Values

SHapley Additive exPlanations

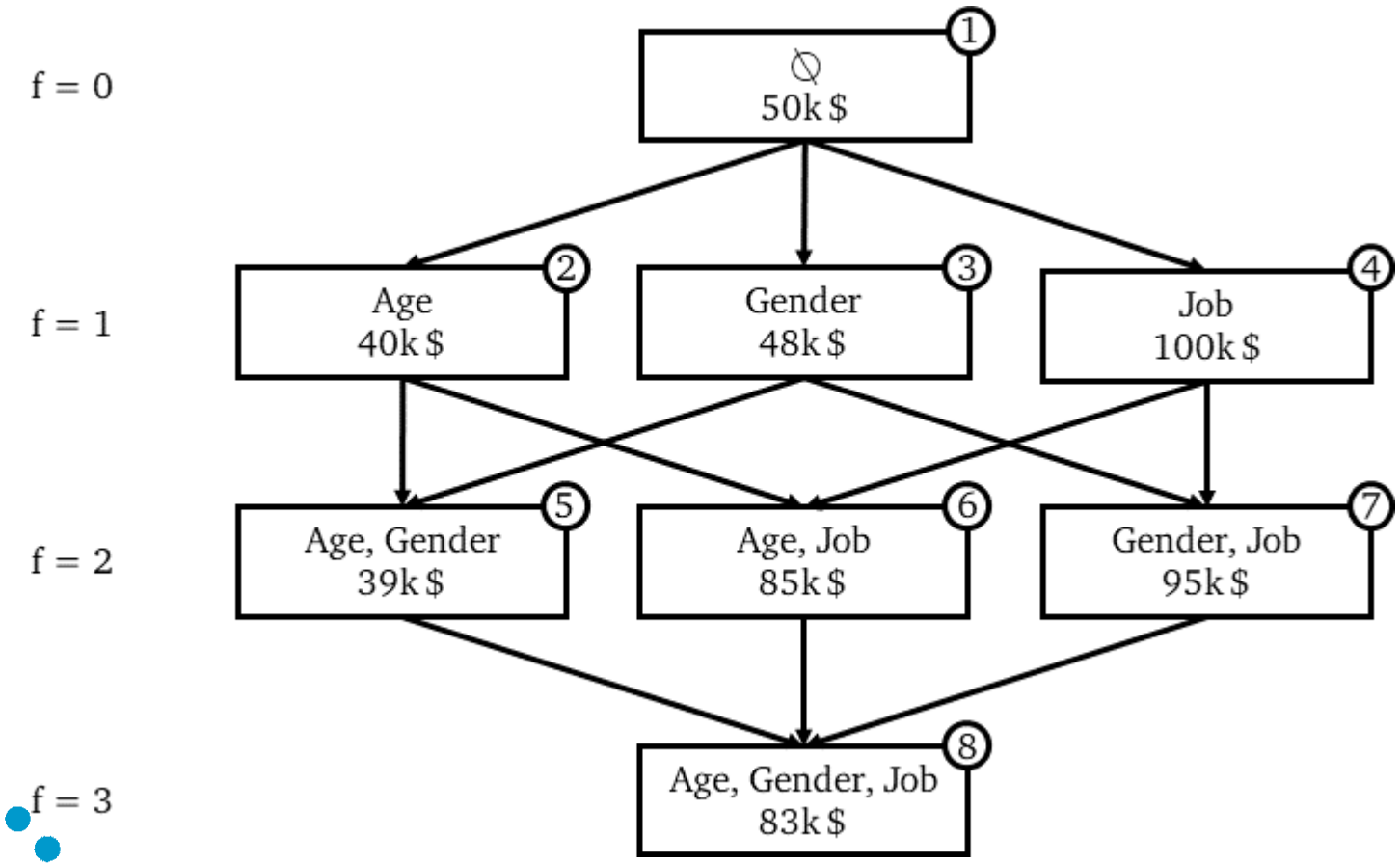
- Can explain any model
- Quantifies the contribution of each feature to the prediction made by the model
- SHAP values explain one prediction/observation
- Considers all combinations of features
- SHAP values interpret the impact of having a certain value for a given feature in comparison to the prediction the model would make if that feature took some baseline value
- Base value – value if all features has their baseline values
- Compare output to base value – which features lead to increase, which to decrease of output

SHapley Additive exPlanations

- Start with a model with no features
- Train models with one feature
- Train models with each combination of two features
- Train models with each combination of up to F features
- Need to train 2^F models

Example

Three features: age, gender, job
Predict: Salary



The model with no features predicts the mean output in the training set – 50k $

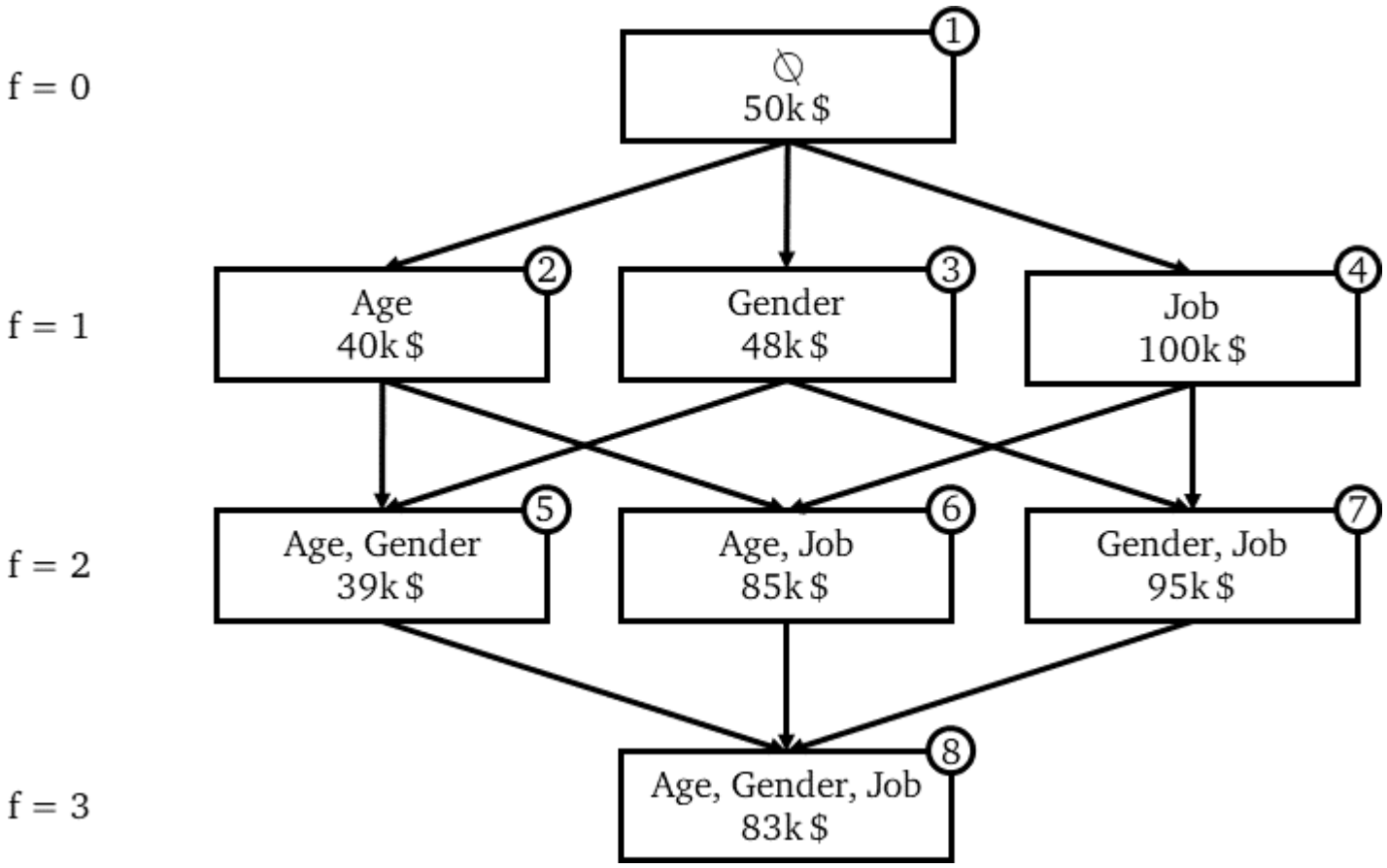
If the age is known, the model's prediction changes to 40k $. Knowing the age lowers the prediction by 10k $ - the marginal contribution of the age feature is -10k $.

If the age and gender are known, the prediction changes to 39k $

If all features are known, the prediction is the 83k $ - this is the actual output of the model and the prediction that we want to explain

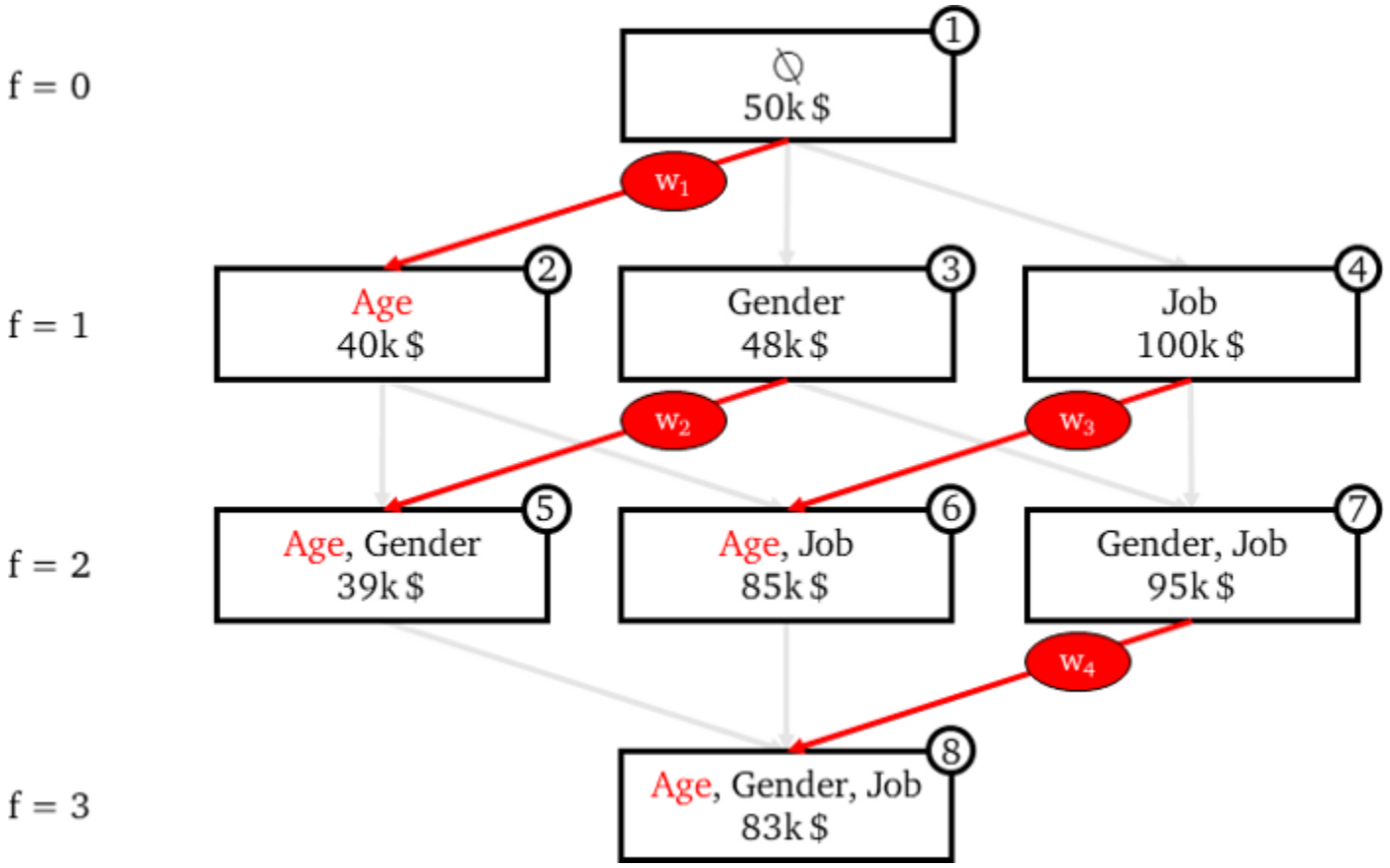
Example

Three features: age, gender, job
Predict: Salary



In order to determine the overall effect of each feature, we need to consider its marginal contribution in all models where it's present.

Example

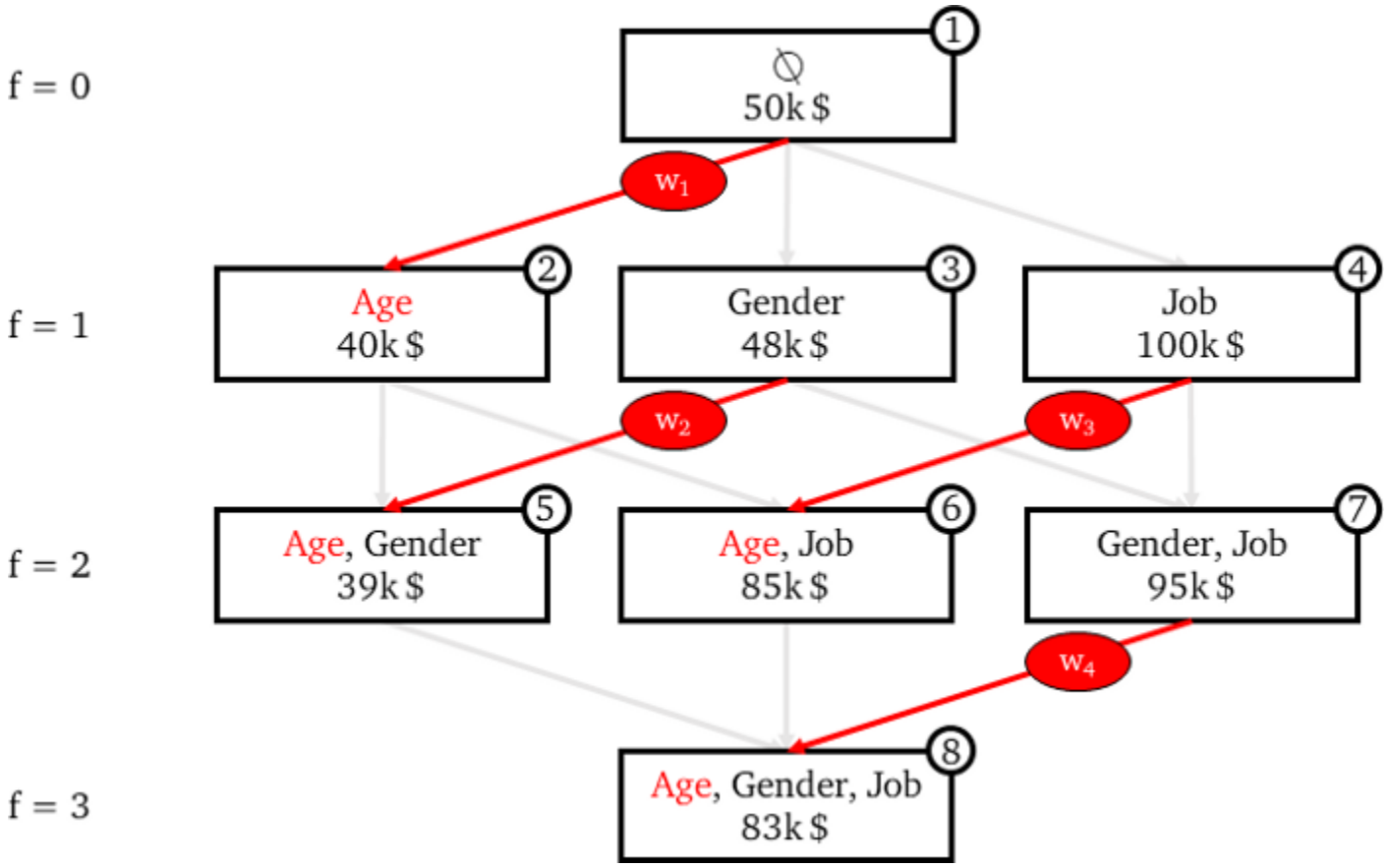


The prediction of the model that uses the features age and gender is lower than the prediction of the model that only uses gender by 9k $. The marginal contribution of age is -9k $.

The prediction of the model that uses the features age and job is lower than the prediction of the model that only uses job by 15k $. The marginal contribution of age is -15k $.

The prediction of the model that uses the all features is lower than the prediction of the model that only uses the age and gender by 12k $. The marginal contribution of age is -12k $.

Example



We multiply each marginal contribution by the weights w_1, w_2, w_3 and w_4.

The weights should sum to 1:

$$w_1 + w_2 + w_3 + w_4 = 1$$

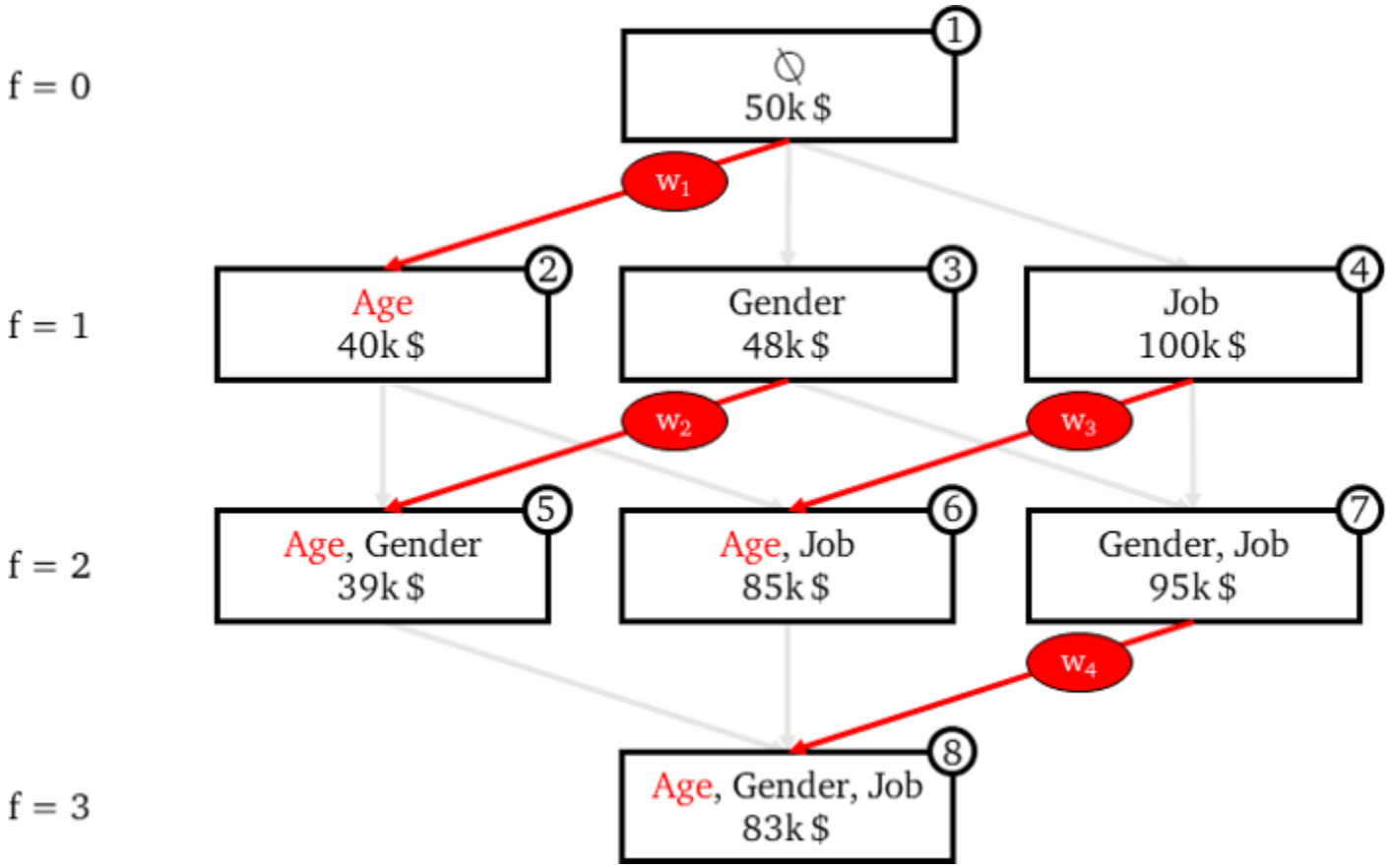
The weights in each row should be equal:

$$w_2 = w_3$$

The sum of all weights in the row should be the same for all rows:

$$w_1 = w_2 + w_3 = w_4$$

Example

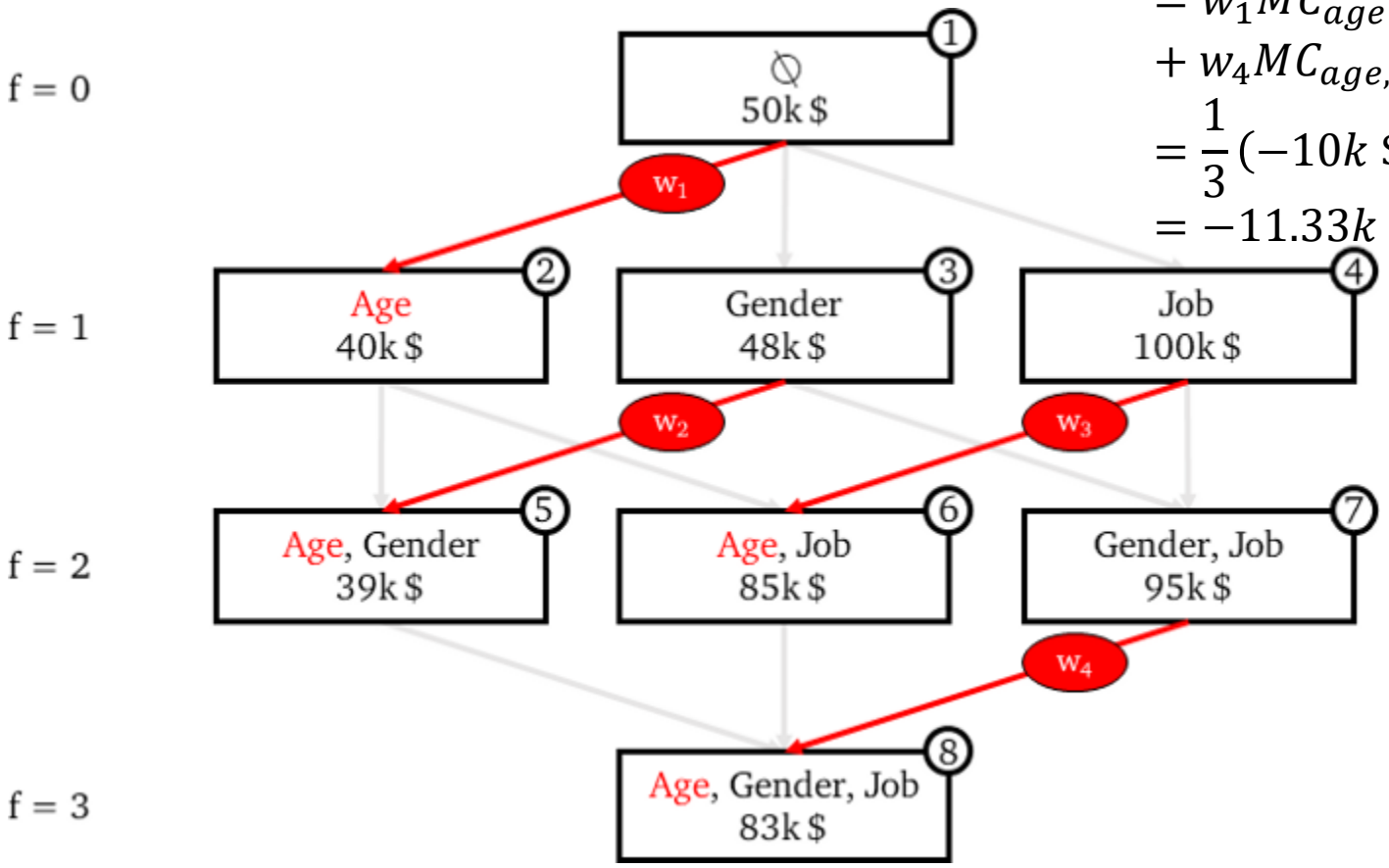


We multiply each marginal contribution by the weights w_1, w_2, w_3 and w_4.

The weights here are:

$$w_1 = \frac{1}{3}$$
$$w_2 = \frac{1}{6}$$
$$w_3 = \frac{1}{6}$$
$$w_4 = \frac{1}{3}$$

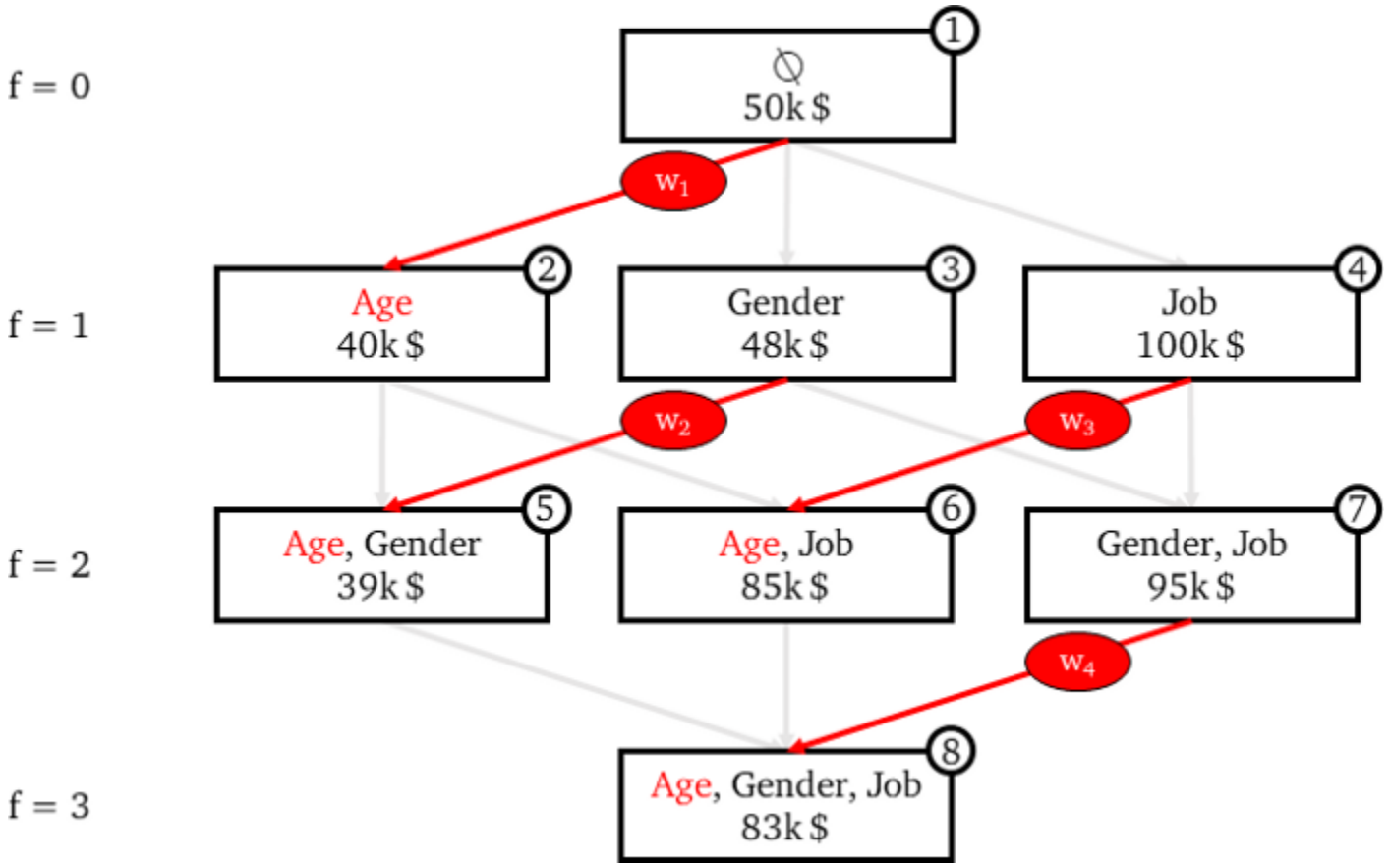
Example



Based on the previous marginal contributions and weights we get:

$$SHAP_{age}$$
$$= w_1 MC_{age} + w_2 MC_{age,gender} + w_3 MC_{age,job}$$
$$+ w_4 MC_{age,gender,job}$$
$$= \frac{1}{3}(-10k\ \$) + \frac{1}{6}(-9k\ \$) + \frac{1}{6}(-15k\ \$) + \frac{1}{3}(-12k\ \$)$$
$$= -11.33k\ \$$$

Example



In order to explain the prediction, we need the SHAP values of all features:

$$SHAP_{age} = -11.33k\ \$$$
$$SHAP_{gender} = -2.33k\ \$$$
$$SHAP_{job} = +46.66k\ \$$$

The sum of all SHAP values is +33k $, which is equal to the difference between the base value (50k $) and the prediction (83k $)

SHAP Phyton library

- Computing the SHAP values by hand isn’t possible for larger models
- The SHAP library for Python can compute SHAP values for all models
- Documentation: https://shap.readthedocs.io/en/latest/index.html

sheldon

smart habitat
for the elderly

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the H2020 Framework Programme
of the European Union

This publication is based upon work from COST Action CA16226: Indoor Living Space Improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

www.**cost**.eu

www.**sheld-on**.eu