

Outline

- 1. Briefly explain the cost-sensitive classification**
- 2. How to generate cost-sensitive output in Weka**
- 3. How to conduct cost-sensitive classification in Weka**

Cost-sensitive classification (CSS)

- **Cost-sensitive classification:** an approach to force machine learning algorithms to consider the costs caused by different kinds of errors. The costs of different kinds of errors are not assumed to be equal, and the objective of CSS is to minimize the expected costs.
- **Basis:** wrongly classifying an instance to a category incurs different costs.

COST	Actually, give a loan	Actually, not give a loan
Should give a loan	0	1
Should not give a loan	5	0

Result 1

a	b	<-- classified as
600	50	a = Good
100	250	b = Bad

Correctly Classified Instances

N = 850

Ratio = 85 %

Cost = $50 \times 1 + 100 \times 5$
= 550

Result 2

a	b	<-- classified as
600	120	a = Good
50	230	b = Bad

Correctly Classified Instances

N = 830

Ratio = 83 %

Cost = $120 \times 1 + 50 \times 5$
= 370

A need for cost-sensitive classifier!