

- Содержание презентации
- Введение в SAS
- Постановка задачи кредитного скоринга
- Основные понятия
- Организация данных
- Организация системы
- Решение поставленной задачи
- Заключение

# Введение в SAS

Зайцев Алексей

МФТИ

## Содержание презентации

- Введение в SAS
- Постановка задачи кредитного скоринга
- Основные понятия
- Организация данных
- Организация системы
- Отчет и визуализация
- Решение поставленной задачи
- Заключение

# Вступление

SAS (Statistical Analysis System) – система разработки со специальным языком, применяемая в data mining и финансовом секторе. Содержит множество независимых пакетов.

## Постановка задачи кредитного скоринга

Задана выборка  $D = \{(y_i, \mathbf{x}_i)\}_{i=1}^n$ . Значение  $y_i \in \{0, 1\}$  – класс объекта, хороший он или плохой заемщик,  $\mathbf{x}_i$  – элемент в пространстве признаков, состоящий из числовых и символьных значений. Задана матрица штрафов  $A$ , она определяет штраф за причисление хорошего заемщика к классу плохих и плохого к классу хороших. Необходимо минимизировать функционал потерь на контрольной выборке:

$$Q(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n_1} \sum_{i=1}^{n_1} a(y_i, \hat{y}_i). \quad (1)$$

Функция  $a(y_i, \hat{y}_i)$  потерь естественным образом определяется матрицей штрафов  $A$ .

В нашем случае матрица штрафов  $A$  задана явно:

$$A = \begin{pmatrix} 0 & 1 \\ 5 & 0 \end{pmatrix} \quad (2)$$

Мы будем решать поставленную задачу классификации методом логистической регрессии.

Содержание презентации  
Введение в SAS  
Постановка задачи кредитного скоринга  
**Основные понятия**  
Организация данных  
Организация системы  
Решение поставленной задачи  
Заключение

DATA и PROC шаги  
Простейшая программа

## Основные понятия

## Основные понятия

- DATA step
  - чтение данных
  - преобразование данных

## Основные понятия

- DATA step
  - чтение данных
  - преобразование данных
- PROC step
  - анализ данных
  - работа с функциями
  - визуализация данных
  - формирование и печать отчетов



## Простейшая программа

```
data gerCred.newData;  
set gerCred.Data;  
run;  
proc print data=gerCred.newData;  
run;
```

## Организация данных

Для записи и хранения данных в SAS используются *библиотеки*. Они бывают временными и постоянными. Обращение к файлу состоит из названия его библиотеки и его собственного названия. Основной тип файлов – *наборы*, состоящие из информации о данных (Data portion) и информации о наборе в целом (Description portion).

## Переменные

Данные в SAS бывают цифровыми либо символьными.

## Переменные

Данные в SAS бывают цифровыми либо символьными. Переменные могут иметь либо символьные, либо цифровые значения. Так же SAS использует свой формат записи дат.

## Пример чтения и вывода данных

```
data gerCred.newData;  
infile 'C:\GermanSAS\german.data-numeric';  
input Variable1-Variable24 Class;  
run;  
proc print data=gerCred.newData;  
var Variable1-Variable6;  
where Variable10>50;  
run;
```

## Пример обработки данных

```
proc sort data=gerCred.newData out=gerCred.sortData;  
by Variable4;  
run;
```

```
title1 'Numeric data for credit score';  
title3 'Selected variables and class';  
proc print data=gerCred.sortData label;  
var Variable10 Variable24 Class;  
label Variable10='age' Variable24='phone';  
format Variable24 phone.;  
run;
```

## Пример анализа данных

```
proc univariate data=gerCred.newData;  
var class;  
run;
```

```
proc corr data=gerCred.newData;  
var Class;  
with Variable1-Variable24;  
run;
```

# Возможные варианты реализации кредитного скоринга в SAS



# Возможные варианты реализации кредитного скоринга в SAS

- SAS STAT

# Возможные варианты реализации кредитного скоринга в SAS

- SAS STAT
- SAS Enterprise

# Возможные варианты реализации кредитного скоринга в SAS

- SAS STAT
- SAS Enterprise
- SAS Enterprise miner

# Возможные варианты реализации кредитного скоринга в SAS

- SAS STAT
- SAS Enterprise
- SAS Enterprise miner
- SAS Credit scoring for Enterprise miner

# Решение в SAS/STAT

```
proc logistic data=gerCred.newData;  
model class=Variable1-Variable24;  
run;
```

- Содержание презентации
- Введение в SAS
- Постановка задачи кредитного скоринга
- Основные понятия
- Организация данных
- Организация системы
- Решение поставленной задачи
- Заключение**

За рамками данного обзора остались многие важные аспекты работы SAS

За рамками данного обзора остались многие важные аспекты работы SAS

- SQL в SAS

За рамками данного обзора остались многие важные аспекты работы SAS

- SQL в SAS
- создание форматов



За рамками данного обзора остались многие важные аспекты работы SAS

- SQL в SAS
- создание форматов
- объединение наборов данных

За рамками данного обзора остались многие важные аспекты работы SAS

- SQL в SAS
- создание форматов
- объединение наборов данных
- ...

- Содержание презентации
- Введение в SAS
- Постановка задачи кредитного скоринга
- Основные понятия
- Организация данных
- Организация системы
- Решение поставленной задачи
- Заключение**

Спасибо за внимание.  
Вопросы?