



POLITECHNIKA WARSZAWSKA  
WYDZIAŁ MATEMATYKI I NAUK INFORMACYJNYCH



PRACA DYPLOMOWA MAGISTERSKA  
NA KIERUNKU MATEMATYKA

**ESTYMACJA W MODELU COXA METODĄ  
STOCHASTYCZNEGO SPADKU GRADIENTU  
DLA  $P \gg N$  NA PRZYKŁADZIE DANYCH  
Z THE CANCER GENOME ATLAS**

AUTOR:  
MARCIN KOSIŃSKI

PROMOTOR:  
PROF. NDZW. DR HAB. INŻ PRZEMYSŁAW BIECEK

WARSZAWA, LIPIEC 2015

.....  
podpis promotora

.....  
podpis autora

# Spis treści

<b>Wprowadzenie</b>	<b>5</b>
<b>1. Estymacja metodą największej wiarygodności</b>	<b>7</b>
<b>2. Numeryczne metody estymacji</b>	<b>9</b>
2.0.1. Ogólne pojęcia związane ze zbieżnością algorytmu . . . . .	9
2.1. Algorytmy spadku wzdłuż gradientu . . . . .	9
2.1.1. Algorytm Cauchy’ego . . . . .	9
2.1.2. Algorytm Raphsona-Newtona . . . . .	9
2.2. Algorytmy stochastycznego spadku wzdłuż gradientu . . . . .	9
2.2.1. Metoda estymacji stochastycznego spadku gradientu I . . . . .	9
2.2.2. Metoda estymacji stochastycznego spadku gradientu II . . . . .	9
<b>3. Model Coxa</b>	<b>11</b>
3.1. Estymacja analityczna w oparciu o metodę największej wiarygodności dla funkcji pseudo/sub-wiarygodności . . . . .	11
3.2. Estymacja numeryczna w oparciu o metodę stochastycznego spadku gradientu rzędu I dla funkcji pseudo/sub-wiarygodności . . . . .	11
<b>4. A gdzie to <math>p \gg n</math>?</b>	<b>13</b>
<b>5. Zaimplementowany algorytm</b>	<b>15</b>
<b>6. Analiza danych genomicznych - model Coxa z estymacją metodą stochastycznego spadku gradientu</b>	<b>17</b>
6.1. Opis i pobranie danych . . . . .	17
6.2. Analiza . . . . .	17
<b>7. Studium przypadku</b>	<b>19</b>
7.1. Porównanie wyników algorytmu SGD z R-N . . . . .	19
<b>A. Kody w R</b>	<b>21</b>
<b>B. Dokumentacja pakietu RTCGA</b>	<b>23</b>



# Wprowadzenie



## Rozdział 1

# Estymacja metodą największej wiarogodności





## Rozdział 2

# Numeryczne metody estymacji

### 2.0.1. Ogólne pojęcia związane ze zbieżnością algorytmu

Warunki stopu itp

### 2.1. Algorytmy spadku wzdłuż gradientu

#### 2.1.1. Algorytm Cauchy’ego

#### 2.1.2. Algorytm Raphsona-Newtona

### 2.2. Algorytmy stochastycznego spadku wzdłuż gradientu

#### 2.2.1. Metoda estymacji stochastycznego spadku gradientu I

Algorytm SGD

#### 2.2.2. Metoda estymacji stochastycznego spadku gradientu II



## Rozdział 3

# Model Coxa

- 3.1. Estymacja analityczna w oparciu o metodę największej wiarogodności dla funkcji pseudo/sub-wiarogodności
- 3.2. Estymacja numeryczna w oparciu o metodę stochastycznego spadku gradientu rzędu I dla funkcji pseudo/sub-wiarogodności



## Rozdział 4

A gdzie to  $p \gg n$ ?



## Rozdział 5

# Zaimplementowany algorytm





## Rozdział 6

# Analiza danych genomicznych - model Coxa z estymacją metodą stochastycznego spadku gradientu

### 6.1. Opis i pobranie danych

### 6.2. Analiza



## Rozdział 7

# Studium przypadku

### 7.1. Porównanie wyników algorytmu SGD z R-N



**Dodatek A**

**Kody w R**



**Dodatek B**

# **Dokumentacja pakietu RTCGA**





Marcin Kosiński  
Nr albumu 123123

Warszawa, 30 kwietnia 2015

## Oświadczenie

Oświadczam, że pracę magisterską pod tytułem „Estymacja w modelu Coxa metodą stochastycznego spadku gradientu dla  $p \gg n$  na przykładzie danych z The Cancer Genome Atlas”, której promotorem jest prof. ndzw. dr hab. inż. Przemysław Biecek wykonałem samodzielnie, co poświadczam własnoręcznym podpisem.

.....  
Marcin Kosiński