

## Ćwiczenie 3

# Liniowe metody aproksymacji funkcji

### Część teoretyczna

Wykład 6: Uczenie się aproksymacji funkcji – modele liniowe.

### Zadania pomocnicze

- Przeanalizuj dokładnie uczenie perceptronu przedstawione na slajdzie 6 wykładu. Zastanów się jak to zaprogramować w Matlabie.
- Zapoznaj się z funkcjami `stepwisefit`, `ridge` i `lasso` (`help Matlab`).

### Zadania do wykonania

Zaprojektuj liniowy aproksymator według algorytmu: (i) perceptronowego, (ii) regresji liniowej, (iii) regresji krokowej, (iv) regresji grzbietowej oraz (v) LASSO.

1. Utwórz zbiór uczący i zbiór testowy – każdy po 100 przykładów 30-wymiarowych:

```

rand('state',nr_gr*r_k);
randn('state',nr_gr*r_k);
%zbiór uczący
x = [ones(100,1) rand(100,30)*2-1 ];
wp = [1 rand(1,10)*0.5+0.5, rand(1,10)*0.3, zeros(1,10)]';
y = (wp'*x' + randn(1,100)*0.1)';
%zbiór testowy
xt = [ones(100,1) rand(100,30)*2-1];
yt = (wp'*xt' + randn(1,100)*0.1)';
    
```

gdzie za `nr_gr` wstaw numer swojej sekcji a za `r_k` aktualny rok kalendarzowy.

`wp` to wektor "prawdziwych" wag. Pierwsza składowa tego wektora to waga  $w_0 = 1$  (wyraz wolny), wagi  $w_1-w_{10}$  przyjmują wartości z zakresu  $[0.5, 1]$ , wagi  $w_{11}-w_{20}$  przyjmują wartości z zakresu  $[0, 0.3]$ , a wagi  $w_{21}-w_{30}$  mają wartość 0.

`randn(1,30)*0.1` to zakłócenie losowe o rozkładzie normalnym ze średnią 0 i odchyleniem standardowym równym 0,1.

Wagi modeli będziemy estymować na zbiorze uczącym. Zbiór testowy służy do wyznaczenia błędu generalizacji (na nowych danych).

2. Przeprowadź uczenie perceptronu (algorytm – wykład). Przyjmij losowe wagi,  $\eta = 0.0002$  i warunek stopu: osiągnięto  $k_{\max} = 1000$  iteracji (epok).
  - Narysuj wykres błędu uczącego w funkcji numeru iteracji.
  - Zanotuj końcową wartość błędu uczącego.
  - Po nauczeniu wyznacz błąd na zbiorze testowym.

Wskazówki:

- losowanie wartości początkowych wag: `w1=rand(31,1);`
- przed uruchomieniem algorytmu należy odwrócić macierze przykładów: `x=x'; xt=xt';`
- w trakcie iteracji należy zapamiętywać błąd, aby sporządzić wykres: `EE(k)=E;`
- aproksymowane wartości funkcji dla zbioru uczącego: `y1=(w1'*x)';`
- aproksymowane wartości funkcji dla zbioru testowego: `y1t=(w1'*xt)';`
- błąd testowy wyznaczamy ze wzoru: `E1t=sum((yt-y1t).^2);`

3. Wyznacz wagi ( $w_2$ ) analitycznie wykorzystując rachunek macierzowy (patrz wykład – model regresji liniowej). Uwaga: powróć do początkowej postaci macierzy przykładów: `x=x'; xt=xt';`

- Wyznacz aproksymowaną wartości funkcji dla zbioru uczącego:  $y_2 = (w_2' * x')'$  ;
- Wyznacz błąd aproksymacji dla zbioru uczącego:  $E_2 = \text{sum}((y - y_2).^2)$  ;
- Analogicznie wyznacz aproksymowaną wartości funkcji  $y_{2t}$  i błąd  $E_{2t}$  dla zbioru testowego.
- Narysuj wykres słupkowy (`bar`) wag prawdziwych  $w_p$ , wag otrzymanych z perceptronu i wag wyznaczonych analitycznie.

4. Użyj regresji krokowej do selekcji istotnych atrybutów (funkcja `stepwisefit`).

- Regresja krokowa:

```
[~,~,~,inmodel]=stepwisefit(x(:,2:end),y);
x3=x(:,[true inmodel]); %usunięcie nieistotnych atrybutów
```

gdzie `inmodel` zawiera jedynki na pozycjach atrybutów istotnych.

- Zwróć uwagę na wyniki wyświetlane w oknie Matlab.
- Podaj numery istotnych atrybutów.
- Wyznacz wagi atrybutów istotnych:  $w_3 = (x_3' * x_3)^{-1} * x_3' * y$ ;
- Wyznacz aproksymowane wartości funkcji dla zbioru uczącego:  $y_3 = (w_3 a' * x_3')'$  ;
- Wyznacz błąd aproksymacji dla zbioru uczącego  $E_3$ .
- Analogicznie wyznacz aproksymowaną wartości funkcji  $y_{3t}$  i błąd  $E_{3t}$  dla zbioru testowego.
- Porównaj wagi  $w_3$  z wagami prawdziwymi, otrzymanymi z perceptronu i wyznaczonymi analitycznie.

5. Użyj regresji grzbietowej do regularyzacji (funkcja `ridge`).

- Wykonaj regresję grzbietową dla  $\lambda = 0:1000$ :

```
lambda=0:1000;
w4=ridge(y,x(:,2:end),lambda,0);
```

`w4` to macierz wag, która w kolejnych kolumnach zawiera wagi dla kolejnych wartości `lambda`.

- Narysuj wagi w zależności od parametru  $\lambda$ .
- Wyznacz aproksymowane wartości funkcji  $y_4$  i błędy  $E_4$  dla zbioru uczącego przy wagach estymowanych dla kolejnych wartości parametru  $\lambda$ :

```
for i=1:length(lambda)
    y4(:,i)=(w4(:,i)'*x')';
    E4(i)=sum((y-y4(:,i)).^2);
end
```

- Sporządź wykres błędów  $E_4$  w zależności od  $\lambda$ .
- Analogicznie wyznacz aproksymowaną wartości funkcji  $y_{4t}$  i błąd  $E_{4t}$  dla zbioru testowego. Sporządź wykres błędów  $E_{4t}$  w zależności od  $\lambda$ .
- Zinterpretuj wykresy i wyniki.

6. Użyj regresji LASSO do regularyzacji (funkcja `lasso`).

- Wykonaj regresję LASSO dla  $\lambda = 0:0.001:1$ :

```
lambda=0:0.001:1;
[w5, FitInfo] = lasso(x(:,2:end),y,'Lambda',lambda);
w5_0=FitInfo.Intercept;
```

$w_5$  to macierz wag (bez wagi dla wyrazu wolnego), która w kolejnych kolumnach zawiera wagi dla kolejnych wartości  $\lambda$ .  $w_{5\_0}$  to wagi dla wyrazu wolnego.

- Narysuj wagi w zależności od parametru  $\lambda$ .
- Wyznacz aproksymowane wartości funkcji  $y_5$  i błędy  $E_5$  dla zbioru uczącego przy wagach estymowanych dla kolejnych wartości parametru  $\lambda$ :

```
for i=1:length(lambda)
    y5(:,i)=([w5_0(i); w5(:,i)]'*x')';
    E5(i)=sum((y-y5(:,i)).^2);
end
```

- Sporządź wykres błędów  $E_5$  w zależności od  $\lambda$ .
- Analogicznie wyznacz aproksymowaną wartości funkcji  $y_{5t}$  i błąd  $E_{5t}$  dla zbioru testowego. Sporządź wykres błędów  $E_{5t}$  w zależności od  $\lambda$ .
- Zinterpretuj wykresy i wyniki.

## Co powinno znaleźć się w sprawozdaniu

- A) Cel ćwiczenia.
- B) Treść zadania.
- C) Opis metod aproksymacji liniowej (nie kopiuj treści wykładu, poszukaj w literaturze i Internecie).
- D) Metodyka rozwiązania – poszczególne instrukcje Matlaba z komentarzem (zachowaj numerację zadań).
- E) Zestawienie wyników (wykresy, tabele z komentarzem).
- F) Wnioski końcowe.

## Zadania dodatkowe dla ambitnych

1. Wykonaj p. 2 i 3 dla danych pobranych od prowadzącego stosując reprezentację rozszerzoną (patrz ostatni slajd wykładu).
2. Zaprogramuj regresję krokową w wersjach z dodawaniem i eliminacją atrybutów (szczegóły uzgodnij z prowadzącym). Zastosuj te algorytmy do danych użytych w ćwiczeniu. Porównaj wyniki.
3. Wykonaj podobne ćwiczenie w innym programie, np. R, Statistica, C#, ... (w uzgodnieniu z prowadzącym).

## Przykładowe zagadnienia i pytania zaliczeniowe

1. Cel i plan ćwiczenia.
2. Materiał ze sprawozdania.
3. Podział metod aproksymacji pod względem sposobu reprezentacji funkcji aproksymowanej.
4. Błąd aproksymacji.
5. Uogólniona reguła delta.
6. Reguła delta dla hipotezy liniowej.
7. Algorytmy uczenia perceptronu.
8. Wagi wyznaczone w modelu regresji liniowej.
9. Idea i tryby regresji krokowej.
10. Idea regresji grzbietowej.
11. Regularyzacja.
12. Idea regresji LASSO.
13. Idea rozszerzonej reprezentacji.

## Do przygotowania na następne zajęcia

1. Zapoznać się z instrukcją do kolejnego ćwiczenia.
2. Zapoznać się z częścią teoretyczną do kolejnego ćwiczenia.

3. Wykonać zadania pomocnicze do kolejnego ćwiczenia.