

Vytěžování dat, cvičení 1: Úvod do Matlabu



Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Fakulta elektrotechnická, ČVUT

Proč probíhá Matlab?

- ▶ Matlab je SW pro vědecko-technické výpočty a de facto průmyslový standard v mnoha odvětvích.
- ▶ Vhodný pro rychlé prototypování a zkoušení aplikací.
- ▶ Obsahuje skriptovací jazyk se spoustou knihoven pro různé oblasti (včetně vytěžování dat).
- ▶ Matlab je v tomto předmětu kompromis mezi "klikacími" nástroji typu RapidMiner a programováním v jazycích typu Java.
- ▶ Dovolí nám hrát si s implementací algoritmů, ale spoustu věcí řeší interně za vás.

Kde získat Matlab?

- ▶ FEL má multilicenci, kterou můžete využít.
- ▶ Kopii instalačního DVD můžete získat na <http://www.fel.cvut.cz/user-info/matlab.html>
- ▶ Zde se musíte přihlásit hlavním přístupovým heslem a pak už můžete stahovat a instalovat.

Základní uživatelské rozhraní

The screenshot displays the MATLAB 7.11.0 (R2010b) user interface. The top menu bar includes File, Edit, Debug, Desktop, Window, and Help. The 'Current Folder' path is shown as `/Users/cepekml/Documents/MATLAB`, which is highlighted with a red box and labeled 'Aktuální pracovní adresář'. The 'Shortcuts' pane on the left shows a file named `factorial.m`, with a red label 'Obsah aktuálního adresáře' pointing to it. The 'Command Window' is the central area, showing a prompt `>>` and the command `x = [1 1 2 2 4 5 6]; mode(x)`. The output is `ans = 1`. A second command `x = [1 1 2 2 4 5 6 6 6]; mode(x)` is shown with a red label 'Příkazový řádek pro zadávání příkazů a jejich výstup'. Below the command window, there is a help section for the `mode` function. The 'Workspace' pane on the right shows a table of variables: A (1x5 double), B (1x5 double), C (1x5 double), ans (1), i (1x3 double), row (1x3 double), sl (1x3 double), str (1x3 double), x (1x9 double), and y (2x2 double). A red label 'Aktuálně definované proměnné a jejich hodnoty (zkuste dvojklik)' points to this table. The 'Command History' pane at the bottom right shows a list of executed commands, with a red label 'Historie příkazů' pointing to it.

File Edit Debug Desktop Window Help

Current Folder: `/Users/cepekml/Documents/MATLAB` Aktuální pracovní adresář

Shortcuts: How to Add What's New

Current Folder: `MATLAB`

factorial.m

Obsah aktuálního adresáře

Command Window

```
>> New to MATLAB? Watch this Video, see Demos, or read Getting Started.
>> x = [1 1 2 2 4 5 6];
>> mode(x)
ans =
     1
>> x = [1 1 2 2 4 5 6 6 6];
>> mode(x)
ans =
     6
```

MODE Return the mode value in time series data

MODE(TS) returns the median of TS.Data

MODE(TS,'PropertyName', PropertyValue,...) includes optional arguments:

- 'MissingData': 'remove' (default) or 'interpolate' indicates how to treat missing data during the calculation
- 'Quality': a vector of integers indicates which quality codes represent missing sample (vector case) or missing observations (>2 dimensional case)
- 'Weighting': 'none' (default) or 'time' When 'time' is used, large time values correspond to 1

See also [timeseries/mean](#), [timeseries/igr](#), [timeseries/std](#)

Workspace

Name	Value	Min	Max
A	<10x5 double>	0.01...	10
B	[1,2,3,4,5,6]	1	6
C	[7,8,9,10,11,12]	7	12
ans	1		6
i	[1,2,3]	1	41...
row	[1,2,3]	1	3
sl	[1,2,3]	1	3
str	[1,2,3]	1	3
x	[1,1,2,2,4,5,6,6,6]	1	6
y	[2,4]	2	4

Command History

```
h(1:ind(a) > 0.5);
fprintf('%d', 5)
fprintf('%d\n', 5)
disp(5)
disp(x)
for i = [1 6 10000 4 3.1415 2.7 1.41]
disp(i); end
factorial(5)
help ce Historie příkazů
A > 0.5
help mode
x = [1 1 2 2 4 5 6];
mode(x)
x = [1 1 2 2 4 5 6 6 6];
mode(x)
```

4 / 24 Úvod do Matlabu

Do příkazového okna zkuste zadat následující příkazy. Jaký je jejich význam a co je výsledkem?

- ▶ $1 + 5$
- ▶ Ukládá se někam výsledek?
- ▶ $x = 3$
- ▶ $y = 1 * 6$
- ▶ $x + y$
- ▶ $z = x + y$
- ▶ $z = x + y;$
- ▶ Jaký je rozdíl mezi předchozími dvěma příkazy?

- ▶ Hlavní síla Matlabu spočívá v práci s maticemi.
- ▶ Jak vypadají matice z následujících příkazů?
- ▶ Jak zjistíme rozměry matic - funkce `shape` `length`

```
A = [1 2; 3 4; 5 6]
```

```
B = [1 2 3; 4 5 6]
```

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

- ▶ Jaké znáte maticové operace?
- ▶ Maticové sčítání, odčítání, násobení, dělení, transpozice
- ▶ `C = [7 8; 9 10; 11 12]`
- ▶ `A * 2`
- ▶ `A + C`
- ▶ `A * B`
- ▶ `A .* C`
- ▶ Jak se liší výsledky posledních dvou příkazů?
- ▶ Co dělá `A'` ?

- ▶ Matice náhodných čísel `rand(<počet prvků v 1. dimenzi>, <v 2. dimenzi>, ...)`
 - ▶ například `A = rand(10, 5)`
- ▶ Velikost matice, počet prvků v jednotlivých dimenzích – `size(A)`
- ▶ Matice samých nul – `zeros(10, 6)`
- ▶ Inverzní matice – `inv(A)`
- ▶ Indikace splnění podmínky – `A > 0.5`
- ▶ Vybere indexy z pole na základě podmínky – `find(A > 0.5)`
- ▶ Spočítá indexy z pole na základě podmínky – `sum(A > 0.5)`

Existuje několik možností získání nápovědy pro příkaz Matlabu.

- ▶ Jednoduchá textová nápověda – příkaz `help <příkaz>`
- ▶ Hypertextová nápověda – příkaz `doc <příkaz>`
- ▶ Mathworld Knowledge Base –
<http://www.mathworks.com/help/techdoc/>

A co vektory?

- ▶ Vektory jsou matice, které mají jen jeden řádek/sloupec.
- ▶ Vytvořte řádkový a sloupcový vektor hodnot 1, 2, 3, 4, 5, 6.
- ▶ `row = [1 2 3 4 5 6]`
- ▶ `s1 = [1; 2; 3; 4; 5; 6]`
- ▶ Jak vytvořím z řádkového vektoru sloupcový?
- ▶ Transpozicí – `row'`

Pokud potřebujete vytvořit posloupnost čísel, můžete využít příslušný operátor ":" (dvojtečka).

Pomocí něj vygeneruje vektor, který obsahuje posloupnost čísel v zadaném rozmezí – zkuste

- ▶ `[1:10]`; `[-5:5]`; `[5.4:15.6]`
- ▶ můžete zadat i krok, se kterým se posloupnost mění. Zkuste: `[1:0.5:10]`; `[5:-1:5]`; `[5.4:0.2:15.6]`

Zopakujte `A = rand(10, 5)`, ať máme všichni stejné rozměry matice.

- ▶ Přístup k jednomu konkrétnímu prvku – `A(1,2)`
- ▶ **POZOR** – indexy se číslují od **1**!
- ▶ Přístup k podmaticím – uvedu rozsahy indexů, které chci v mít podmatici `A(3:5, 1:3)`
- ▶ Místo čísel můžu uvést i vektory indexů, které chci zobrazit.
- ▶ Zkuste zobrazit prvky ve řádcích 1., 5., 3., 8. a sloupcích 3., 2., 1.
- ▶ `x = [1 5 3 8]; y = [3 2 1]; A(x,y)`

Přiřazování hodnot do matice

- ▶ Úplně stejně jako výběr prvků na minulém slajdu, jen výběr umístím na levou stranu přiřazení.
- ▶ Přiřazení jedné hodnoty – přiřaďte 10 do levého horního rohu matice.
- ▶ $A(1,1) = 10$
- ▶ Přiřazení do podmatice – přiřaďte hodnoty 1, 2, 3, 4 na souřadnice (2,2), (2,4), (4,2), (4,4)
- ▶ $x=[2 \ 4]$; $y=[2 \ 4]$; $A(x,y) = [1 \ 3; 2 \ 4]$

- ▶ V Matlabu, stejně jako v jiných skriptovacích jazycích, proměnné nemají pevný datový typ.
- ▶ Základní datové typy jsou:
 - ▶ Čísla, Logické hodnoty, Řetězce
 - ▶ Matice – matice hodnot jednoho datového typu
 - ▶ Struktury – skupina několika pojmenovaných hodnot zabalených do jedné proměnné
 - ▶ Buňková pole (Cell arrays) – pole hodnotu různých datových typů
 - ▶ Odkazy (Handles)
 - ▶ Objekty

Více o datových typech se lze dozvědět na http://www.mathworks.com/help/techdoc/matlab_prog/f2-43934.html

Doteď jsme zkoušeli interaktivní práci se systémem Matlab. Teď zkusíme programování.

Programy se zadávají do tzv. M-souborů (M-file) což je jen textový soubor s koncovkou `.m`. Existují dva typy M-souborů

- ▶ Skripty – obsahují jen posloupnost příkazů Matlabu,
- ▶ Soubory definující funkci – obsahuje definici funkce, kterou lze využít při interaktivní práci s Matlabem nebo v jiných M-souborech.

Pro editaci obou typů M-souborů můžete použít:

- ▶ libovolný textový editor,
- ▶ editor integrovaný přímo do Matlabu.

My použijeme druhou možnost. Interní editor se spouští buď příkazem `edit` nebo z menu `File > New Script` resp. `File > New Function`.

Základní struktura funkce vypadá takto:

```
function [soucet,rozdil] = SlozitaFunkce(a,b)
%
% Funkce se jmenuje SlozitaFunkce a ma dva vstupni
% parametry - a, b. A dva vystupni parametry -
% soucet, rozdil.
% Tohle je help k funkci SlozitaFunkce. Vypisuje se
% zadanim prikazu 'help SlozitaFunkce'.
%
%Naplneni vystupniho parametru se deje prirazenim
    soucet = a+b;
    rozdil = a-b;
end % Nepovinne end
```

Funkce **musí** být uložena v souboru slozitafunkce.m.

- ▶ Funkci můžete zavolat pomocí jejího jména.
- ▶ `[s, r] = slozitafunkce(10, 4)`
- ▶ Jak Matlab zjistí, kterou funkci voláte?
 - ▶ Matlab se podívá do aktuálního adresáře, zda v aktuálním adresáři existuje soubor `slozitafunkce.m`
 - ▶ Pokud Matlab funkci nenajde v aktuálním adresáři, podívá se na vnitřní proměnné `PATH` na seznam adresářů, které se mají prohledávat a zkusí, zda některý z nich funkci neobsahuje.
- ▶ Z toho plyne, že je důležité, ve kterém adresáři se nacházíte :).

Provádění funkce můžete ovlivnit pomocí standardních konstrukcí, které znáte z jiných programovacích jazyků.

- ▶ Podmínky
 - ▶ if – then – else
 - ▶ switch – case
- ▶ Cykly
 - ▶ for
 - ▶ while

if – then – else

```
if x > 11
    disp('x je vetsi nez 11.');
```

elseif x < 5

```
    disp('x je mensi nez 11 a take mensi nez 5.')
```

else

```
    disp('Je to nejak uplne jinak.');
```

end

switch – case

```
switch x
    case {0, 1, 2, 3}
        disp('x je v intervalu 0-3.');
```

case 4

```
        disp('x je 4.');
```

otherwise

```
        disp('x je uplne jine.')
```

end

For cyklus funguje trochu jinak než jste zvyklí z Javy. For cyklus v Matlabu iteruje přes všechny hodnoty vektoru.

```
for i = 1:10
    fprintf('%d ',i); %muzete pouzit take disp(i)
end
```

```
for i = [1 6 10000 4 3.1415 2.7 1.41]
    disp(i);
end
```

POZOR – použití cyklů je extrémně nevýhodné, protože Matlab parsuje každý řádek znovu ⇒ pokud můžete zkuste cykly obejít (například maticovými operacemi, příkazy `find`, `sum`, `arrayfun`...).

While cyklus je mnohem blíží tomu, co znáte z jiných jazyků.

```
i = 1;
while i < 10
    fprintf('%d',i);
    i = i+1;
end
```

Existují i příkazy na přerušení cyklů

- ▶ break – ukončuje cyklus if $i = 5$, break, end
- ▶ continue – přeskakuje zbytek iterace if $i = 5$, continue, end

Příklad - faktorial

```
function fac = factorial(n)
if n < 0
    disp('n musi byt vetsi nez 0!');
    fac = 0;
    return;
end
if n == 0
    fac = 1;
    return;
end
fac = 1;
for i = 1:n
    fac = fac * i;
end
end
```

Zde jsou odkazy na další výukové materiály o Matlabu

- ▶ <http://labe.felk.cvut.cz/~posik/y33aui/uvod-do-matlabu/>
- ▶ http://www.mathworks.com/help/techdoc/matlab_product_page.html
- ▶ <http://www.mathworks.com/moler/intro.pdf>
- ▶ <http://www.maths.dundee.ac.uk/~ftp/na-reports/MatlabNotes.pdf>
- ▶ a mnoho dalších...