

## ÚLOHA 4

### Logistická regrese

*Zadáno na cvičení: 5*  
*Mezní termín: 8.11. 2017*  
*Maximální počet bodů: 15*  
**Povinná úloha**

### Zadání

Stáhněte si archiv *logReg.zip* ze stránky 3 *Logistická regrese*. Archiv obsahuje tyto soubory:

- *crossValidation.m* – křížová validace pro využití stejných dat pro trénování i testování
- *data1.txt* – vstupní data pro první část
- *data2.mat* – vstupní data pro druhou část v binárním formátu MATLABu.
- *displayData.m* – vizualizace dat
- *featureNormalize.m*<sup>0</sup> – škálování příznaků
- *fmincg.m* – pokročilá optimalizační funkce
- *getLogisticRegression.m* – pro vytvoření logistické regrese.
- *gradientDescent.m*<sup>0</sup> – gradientní sestup
- *logRegCost.m*<sup>1</sup> – cenová funkce logistické regrese
- *logRegExample.m* – spouštěcí skript první části
- *logRegMultiExample.m* – spouštěcí skript druhé části
- *logRegPredict.m*<sup>1</sup> – hypotéza logistické regrese
- *oneVsAll.m*<sup>2</sup> – trénovací funkce pro vytvoření klasifikátoru do více tříd technikou one-vs-all.
- *plotData.m* – vizualizace dat
- *plotDecisionBoundary*<sup>1</sup> – Vizualizace rozhodovací hranice
- *predictOneVsAll.m*<sup>2</sup> – Hypotéza pro techniku one-vs-all
- *sigmoid.m*<sup>1</sup> – logistická funkce (sidmoida)

Soubory označené <sup>0</sup> máte naprogramované z předchozí úlohy.

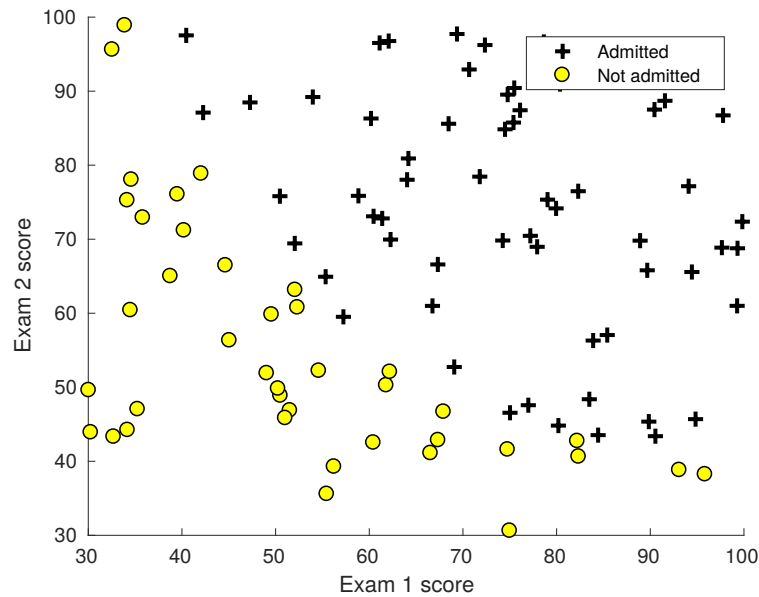
Soubory označené <sup>1</sup> budete doplňovat v první části.

Soubory označené <sup>2</sup> budete doplňovat ve druhé části.

## 1 Klasifikace do dvou tříd

### Vstupní data

Máme k dispozici výsledky studentů ve dvou srovnávacích testech a predikujeme, jestli budou nebo nebudou přijati (přesněji s jakou pravděpodobností budou přijati) na univerzitu.



Obrázek 1: Vizualizace dat.

## Úkoly

### 1. Naprogramujte cenovou funkci a hypotézu.

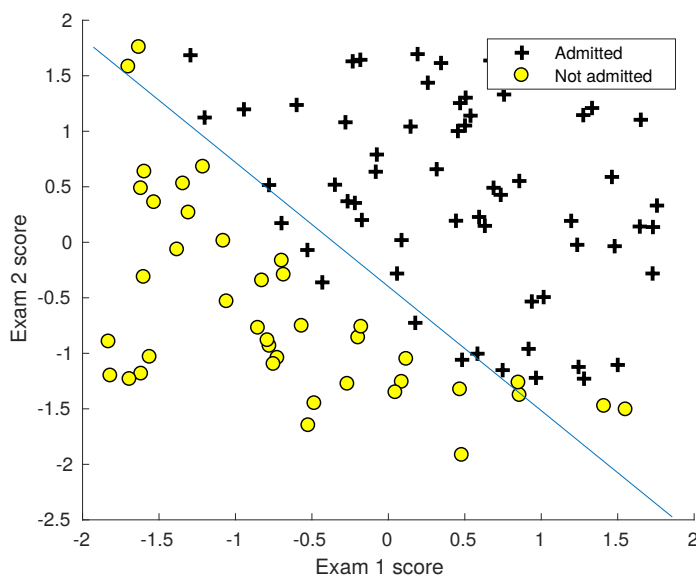
Soubory *sigmoid.m*, *logRegPredict.m* a *logRegCost.m*.

### 2. V souboru *train.m* inicializujte a vyladte parametry gradientního sestupu tak, abyste se dostali do optima (podle vizualizace chyby v závislosti na parametrech theta a počtu iterací, viz předchozí úlohy).

### 3. V souboru *plotDecisionBoundary.m* naprogramujte vykreslení rozhodovací hranice. Rozhodovací hranice je přímka, takže k jejímu vykreslení stačí dva body. Zvolte si $x_1$ pro dva body a dopočítejte pro ně $x_2$ .

Očekávaný výsledek po tomto kroku můžete vidět na obrázku 2.

### 4. Vypočítejte pravděpodobnost, že student s hodnocením 45 v prvním testu a 85 ve druhém bude přijat.



Obrázek 2: Rozhodovací hranice

## 2 Klasifikace do více tříd

V této části použijte logistickou regresi k rozpoznávání ručně psaných číslic. Čísla jsou uložena jako bitmapa 20x20. Přímou v této surové podobě vstupují do klasifikátoru, což rozhodně není ideální reprezentace, ale pro demonstrační účely postačí.



Obrázek 3: Část trénovacích dat

### Úkoly

1. **Naprogramujte obecný one-vs-all klasifikátor.** Budete doplňovat soubory *oneVsAll.m* a *predictOneVsAll.m*.

Program vypíše přesnost na trénovacích a testovacích datech (cca 88% a 87%).