
ÚLOHA 6

Regularizace

Zadáno na cvičení: 10
Mezní termín: 15.12. 2022
Maximální počet bodů: 10
Nepovinná úloha

Zadání

Struktura kódu je následující:

- *data1.txt* – vstupní data pro první část úlohy
- *data2.txt* – vstupní data pro druhou část úlohy
- *model/LinearRegression*⁰★ – třída implementující celou funkcionalitu lineární regrese (hypotéza, pokutová funkce, gradient)
- *model/LogisticRegression*⁰★ – třída implementující celou funkcionalitu logistické regrese (hypotéza, pokutová funkce, gradient)
- *optimize/Optimizer* – generický optimalizační algoritmus
- *optimize/GradientDescent*⁰ – gradientní sestup.
- *visualize.py* – kontrolní vizualizace podobné jako v předchozí úloze
- *regularization.py* – Hlavní skript celé aplikace
- *regularization/learning_curve*()★ – příprava dat pro vizualizaci učící křivky
- *regularization/validation_curve*()★ – příprava dat pro vizualizaci validační křivky

Třídy/funkce označené ★ budete doplňovat.

Třídy/funkce označené ⁰ máte implementované z předchozích úloh

1 Regularizovaná regrese

1. Doplňte regularizační člen do cenové funkce lineární a logistické regrese a do normální rovnice.

2 Přeučení a nedoučení

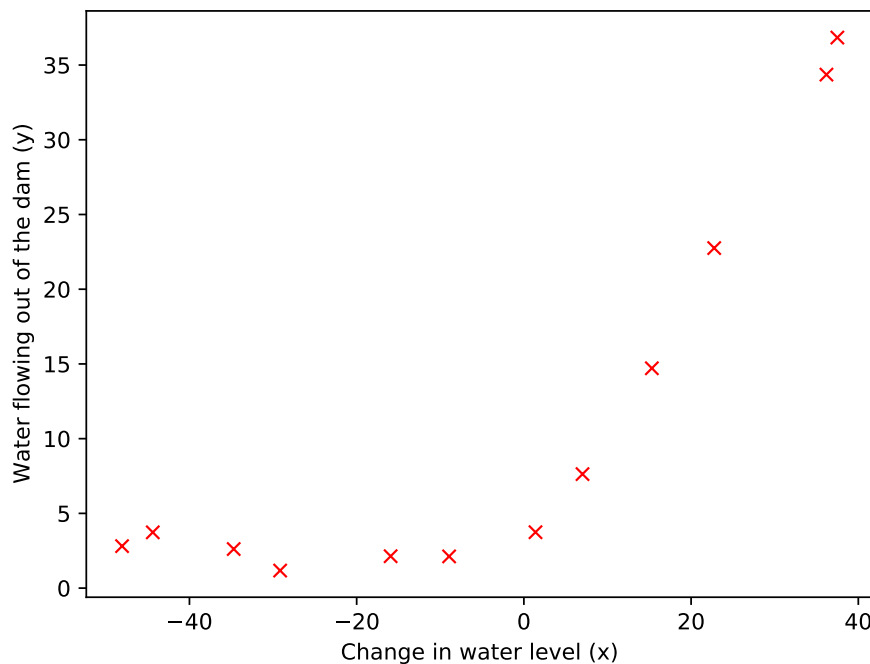
První úloha demonstruje problém přeučení a nedoučení na jednoduché polynomiální regresi.

Vstupní data

Máme k dispozici historické údaje o závislosti množství vody, které vyteče z přehrady na zvýšení hladiny vody.

Data jsou rozdělená na tři množiny:

1. **Trénovací data** – Využívají se k natrénování klasifikátoru.
2. **Validační data** – Využívají se k ladění hyperparametrů (alfa, lambda ...).
3. **Testovací data** – Využívají se pro vyhodnocení úspěšnosti klasifikátoru.



Obrázek 1: Vizualizace dat.

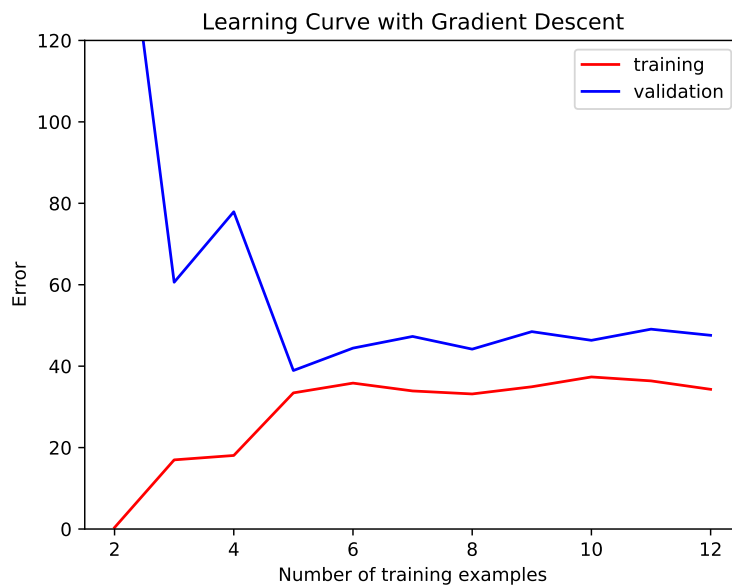
Úkoly

1. Naprogramujte funkci *learningCurve* pro výpočet dat pro učící křivku.

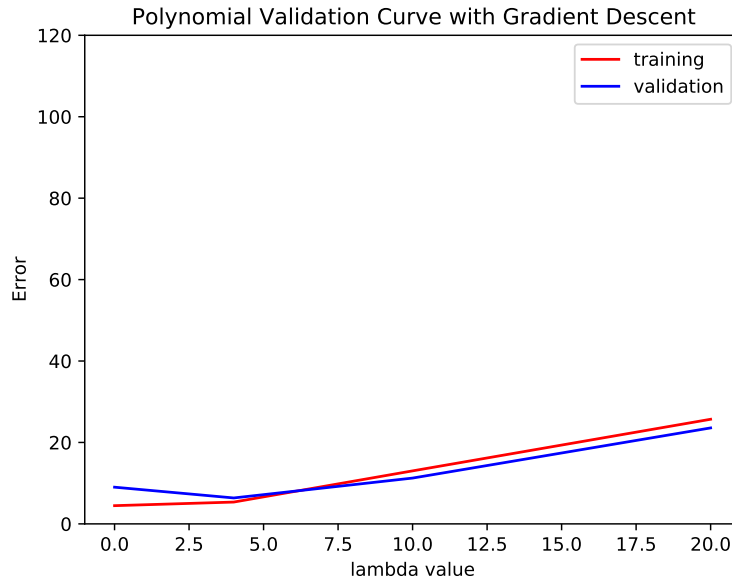
Učící křivka je závislost chyby (například hodnoty cenové funkce) trénovacích a validačních dat na počtu **trénovacích** dat.

2. Naprogramujte funkci *validationCurve* pro výpočet dat pro validační křivku.

Validační křivka je závislost chyby trénovacích a validačních dat na regularizačním koeficientu λ a slouží k nalezení optimální hodnoty λ .



Obrázek 2: Učící křivka

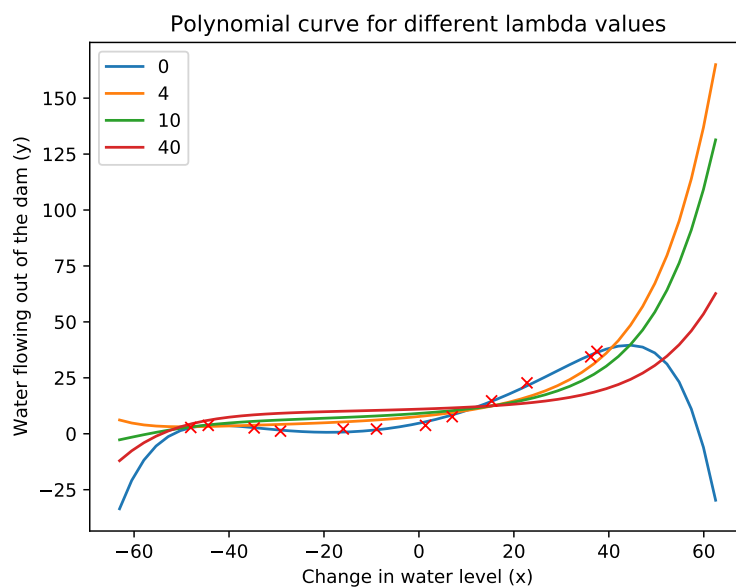


Obrázek 3: Validační křivka

3. Vyzkoušejte si, co s učící křivkou dělá změna hodnoty λ .

Nemusíte se dostat k nějakému výsledku, jen se zamyslete, co vizualizace vlastně říkájí o průběhu učení.

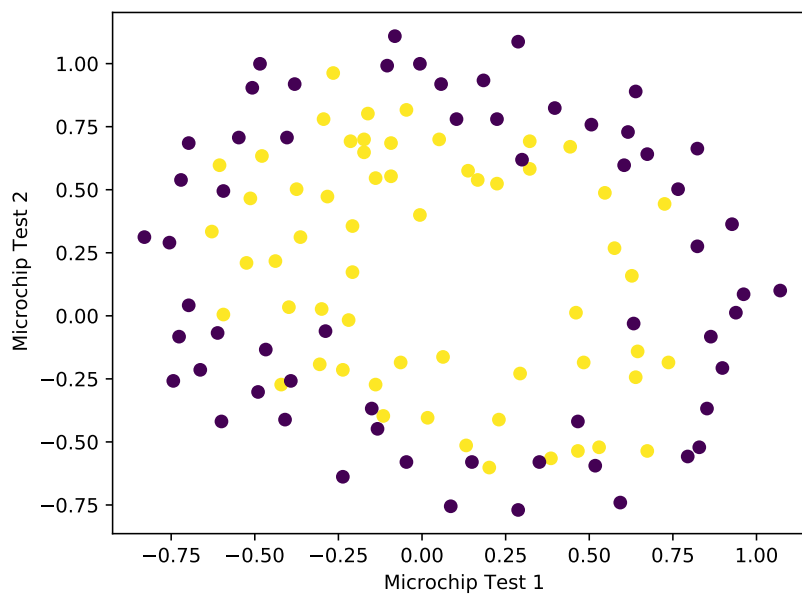
4. Pokuste se zdůvodnit průběhy chyb v jednotlivých grafech. Stačí krátký komentář k jednotlivým grafům.



Obrázek 4: hypotéza pro různé hodnoty λ

3 Polynomiální klasifikace

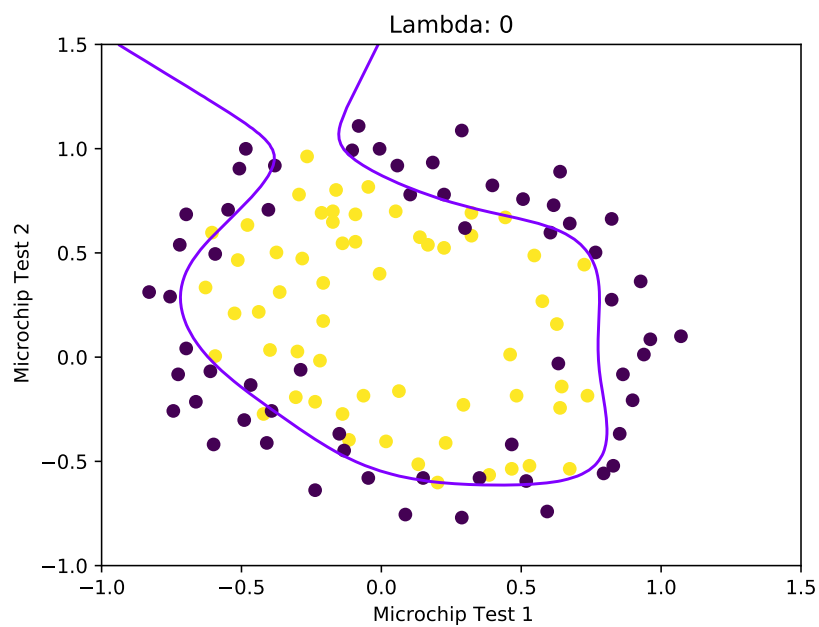
V této části pouze pustíme algoritmus regularizované logistické regrese. Testovací skript používá polynomiální logistickou regresi na klasifikaci lineárně neseparabilních dat.



Obrázek 5: Vizualizace dat.

Úkoly

1. Vykreslete rozhodovací hranici pro několik různých hodnot λ .



Obrázek 6: Rozhodovací hranice bez regularizace