

Úkol A – Explorace dat a lineární regrese – Forest fires

ZS 2017

Cíl

Cílem prvního úkolu je seznámit se s daty o lesních požárech v severovýchodním Portugalsku a pomocí lineární regrese zjistit, jaký je vztah mezi plochou spáleniště a lokálním počasím.

Data

Lesní požáry je událost spojená s velkými riziky pro lesní ekonomii a současně ohrožující lidi, i když z hlediska ekologie a životního prostředí se jedná o přirozenou formu obnovy lesních společenství. Zvládnutelnost lesního požáru a jeho následky jsou ovlivněny počasím v době požáru. Data z portugalské studie nabízí zajímavé informace o vztazích mezi jednotlivými parametry počasí a velikostí lesních požárů.

Explorace dat

Data z přírodního parku Montesinho v Portugalsku obsahují 13 příznaků, jejich stručný popis je uveden v tabulce 1 níže. V rámci explorace dat prozkoumejte příznaky týkající se velikosti spáleniště a počasí (plocha spáleniště, teplota, relativní vlhkost, rychlost větru, dešťové srážky). Seznamte s typem dat v jednotlivých příznacích, rozsahem jejich hodnot, jejich rozdělením, případně počtem chybějících a odlehlých hodnot. Zvažte, jestli některé z příznaků není potřeba transformovat pro potřeby dalších analýz. Prozkoumejte závislosti mezi jednotlivými příznaky a zajímavé závislosti vizualizujte.

Modelování pomocí lineární regrese

Vlivy počasí na požár se zdají být poměrně přímočaré – vítr oheň rozfouká, déšť uhasí, apod. Na základě exploračních vizualizací a modelů, rozhodněte které parametry počasí jsou pro polchu spáleniště vhodné prediktory a modelujte vztah mezi vybranými parametry a plochou spáleniště pomocí lineárního modelu. Výsledný model zhodnoťte z hlediska přesnosti modelu a jeho statistické významnosti.

Požadované kroky analýzy

Průzkumová analýza. [4 body]

- Dimenze dat (počet příznaků, počet instancí)
- Chybějící a odlehlé hodnoty, transformace dat
- Vizualizace

Modelování vztahu mezi plochou spáleniště a parametry počasí [6 bodů]

- Formální zápis modelu
- Interpretace koeficientů modelu
- Přesnost modelu
- Statistická významnost

Výsledky upravte do formy zprávy, která bude obsahovat stručný **úvod**, popis metod, které jste použili, v sekci **metody**, výsledky jejich aplikace na data v sekci **výsledky** a závěry, které jste zjistili interpretací výsledků v sekci **závěr**. Maximální délka zprávy je 3 stránky. Zprávu ve formátu pdf odevzdejte pomocí UploadSystemu.

X	odvozená x-ová souřadnice (od 1 do 9)
Y	odvozená y-ová souřadnice (od 1 do 9)
month	měsíc v roce (leden až prosinec)
day	den v týdnu (pondělí až neděle)
FFMC	FFMC kód
DMC	DMC kód
DC	DC kód
ISI	ISI index
temp	venkovní teplota (ve °C)
RH	venkovní relativní vlhkost (v %)
wind	venkovní rychlost větru (v km/h)
rain	venovní déšť (v mm/m ²)
area	celková plocha spáleniště (v ha)

Tabulka 1: Popis proměnných z studie lesních požárů v severovýchodním Portugalsku

Bonusová úloha – rozdíly v závažnosti lesních požárů v závislosti na roční době

Lineární regrese, nebo obecněji lineární model, je možné k testování řady poměrně složitých hypotéz. Lineární model může sloužit k porovnání závislostí na základě kategorických proměnných. Zajímavým příkladem v tomto může být vztah mezi roční dobou a vlivem počasí na závažnost lesních požárů. Lineární model lze použít k takovému porovnání. Lze na základě dat ze studie lesních požárů zjistit, jestli je v některých měsících v roce vliv některých parametrů počasí významně závažnější?

Bonusová úloha umožňuje získání až 4 dalších bodů, ale vyžaduje samostudium teorie a praktického použití lineárních modelů nad rámec cvičení DVZ.