Project 1: Stack Overflow Data

Columns of interest:

- Employment Status ← How did people get hired
- Pronounce "GIF"
- Job Seeking Status
- Hours Per Week
- Assess... (what jobs people live)
- Circle keys
- Important Hiring (what is important when hiring...)
- Tabs spaces
- Influence (contributing outcomes at company)

↳ Aim, continue where ML algorithm paid off?

↳ create estimater for salary?

→ Data Mining Approach to predict Forest Fires using Meteorological Data

- → Abs.
- → Intro.
- → Forest Fire Dat
- → Data mining Models
- → Experemental Results
- → Conclusions

Keywords

- → SVM
- → R. forrest
- → feature selection
- → Regression
- → Area
- → MAP
- → RMSE
- → NNs
- → R
- → Fire weather index
- → Montesinho

Abstract

Forrest fires
↳ expensive (lives, damage)
→ try to prevent using meteorological measurements

→ predicted using SVM
↳ can we predict using XAI?

Conclusions

→ SVM is capable of predicting small fires
↳ lower predictive power for large fires

Introduction

Fires ← caused by human negligence / lightning
↳ Damage reduced by fast response
↳ Quick data gathering
- satelite ✗ Bad resolution & expensive
- IR ✗ high costs
- Smoke

Forest Fire Data

FWI ← rates fire danger

- FFMC - moisture content surf. litter
  fire ignition & spread
- DMC & DC - moisture of shallow & deep org. layers
- ISI - fire velocity spread
- BUI - available fuel

} indicator of fire intensity

Park

↳ High flora & fauna ~~density~~ diversity

avg temp = 5° - 12°

↳ data collected between 2000 & 2003
↳ could be a problem

↳ coordinates in 9x9 grid !!

→ Day of wk important because fire has human causes.

↳ Data has already been cleaned

Data Mining Models

$$MAD = \frac{\Sigma |y - y_i|}{N}, \quad RMSE = \sqrt{\frac{\Sigma (y - y_i)^2}{N}}$$

↳ sensitive to high error

→ Most of burned f(area)

↳ $f(area) = \ln(area + 1)$

↖ log(0) is impossible

Models considered:

- multiple reg
- regression trees
- random reg. forest
- NN
- SVM *

Experimental Results

→ preprocessing

↳ nominal categorical: 1-of-C encoding

↳ cont. vars normalized to $N(0, 1)$

SVM, fit using seq. min. optimization algorithm

↑ hyper-parameters found using grid search