

MAD 3 projekt
Regrese koncentrace pevných látek v Pekingu

Bc. Moravec Vojtěch

ZS 2019/2020

Obsah

1	Popis problému	3
2	Popis datasetu	3
3	Předzpracování	3
4	Regrese	3

1 Popis problému

Námi vybraný dataset [1] se zaměřuje na znečištění vzduchu v Pekingu od 1. Ledna 2010 do 31. Prosince 2014. Tento dataset jsme získali z UCI Machine Learning Repository [2]. Znečištěním rozumíme koncentraci v mikrogramech na metr krychlový ($\mu\text{g}/\text{m}^3$) pevných částic ve vzduchu. V našem případě se jedná o částice $\text{PM}_{2.5}$, jejichž průměr je maximálně $2.5\text{ }\mu\text{m}$. Naším cílem je tedy provést explorační analýzu datasetu a následně provést předpověď, regresi koncentrace znečištění vzduchu vzhledem k času a přiloženým meteorologickým datům.

2 Popis datasetu

Dataset obsahuje celkem 43 824 záznamů a 12 atributů, nepočítáme-li číslo řádku. Ve 2 067 řádcích chybí cílová koncentrace a proto byly tyto řádky ihned odstraněny. Všechny záznamy jsou snímány v čase, tedy známe datum a čas měření, z něhož jsme vytvořili další atribut den v týdnu. Pro regresi máme tedy k dispozici 12 atributů, kde většina je numerická, až na směr větru resp. den v týdnu, které nabývají 4 resp. 7 hodnot.

3 Předzpracování

4 Regrese

Reference

- [1] Liang X., Zou T., Guo B., Li S., Zhang H., Zhang S., Huang H., and Chen S. X., “Assessing beijing’s pm2.5 pollution: severity, weather impact, apec and winter heating,” 2015.
- [2] D. Dua and C. Graff, “UCI machine learning repository,” 2017.