

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA BIOMEDICÍNSKÉHO INŽENÝRSTVÍ
Katedra biomedicínské informatiky



Diabetes indiani

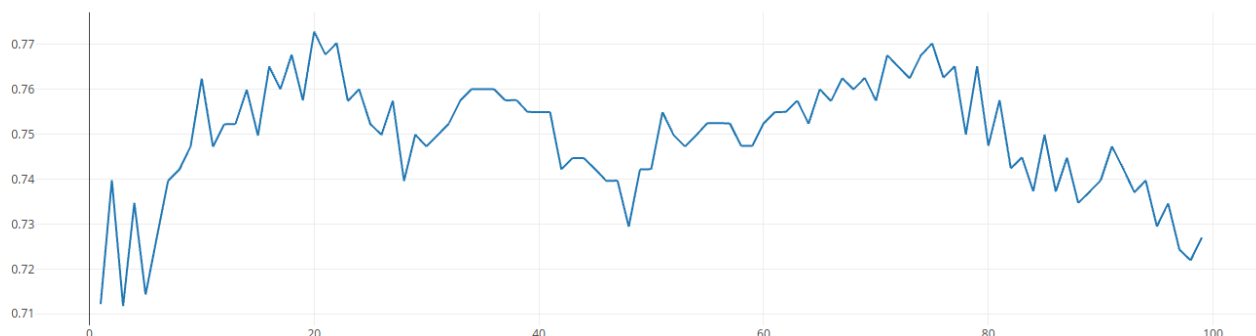
Úkol C

Autor práce: Karavaev Aleksei

květen 2020

2.2 Úspěšnost klasifikátorů

U SVC a Logistické regresi je hodnota přesnosti cca 0.78. Přesnost KNC závisí na vstupním parametru "počet nejbližších sousedů". Nejlepší přesnosti model dosahuje při počtu 20. Výsledný "voting" model má přesnost 0.79.



Obrázek 2.2: Přesnost KNC VS počet sousedů

2.3 Omezení příznaků

Nejvíce ovlivňují rozhodnutí glukosa, age a insulin. Což dává smysl, protože diabetes je poruchou metabolismu cukru. Pokud se model udělá jenom na základě těchto příznaků přesnost bude 0.78, což je srovnatelný s plným modelem.

2.4 Očekávaná úspěšnost

Očekávaná úspěšnost v případě nově příchozího pacienta by byla 0.79.

2.5 Pravděpodobnost nemocného pacienta

Počet nemocných v datasetu je 130, celkový počet je 392. To znamená, že pravděpodobnost že nově příchozí člověk bude nemocný je 0.33.

3 Výsledky

3.1 Zdravý jako nemocný a nemocný jako zdravý

Počet špatně klasifikovaných zdravých je 19, nemocných - 54. To znamená že pravděpodobnost, že zdravý člověk bude klasifikován jako nemocný je $19/262 = 0.072$, nemocný jako zdravý - $54/130 = 0.41$.

3.2 Vyvažování množin

Ano má. Nemocné lide jsou klasifikované jako zdravé častěji, než naopak. Což je ale nežadoucí v případě, když chceme udělat diagnózu na základě modelu. Mnohem důležitější je správně klasifikovat nemocného pacienta. Z tohoto důvodu jsem rozhodl pro vyvážení datasetu. Po vyvážení pravděpodobnosti staly 0.22 pro chybu klasifikaci zdravého jako nemocného a 0.26 pro klasifikaci nemocného jako zdravého.

Seznam obrázků

2.1	Výběr klasifikátoru	2
2.2	Přesnost KNC VS počet sousedů	3