**Strojové učenie a neurónové siete**

**Zadanie číslo 3**

Andrej Dubeň

74657

Obsah

[Zadanie 2](#_Toc498070795)

[Riešenie 3](#_Toc498070796)

[MLP 3](#_Toc498070797)

[SVM 5](#_Toc498070798)

# Zadanie

V tomto zadaní bude našou úlohou identifikovať znak, ktorý sa nachádza na obrázku. Túto úlohu budeme riešiť pomocou klasifikátorov ako SVM (support vector machine) a MLP (multi layer perceptron). Ako vstup pre natrénovanie našich klasifikátorov budeme využívať trénovaciu množinu pripravenú v predchádzajúcom zadaní. Na natrénovanie našich klasifikátorov vyberte 50, 100, 1000, 5000, 20000 vzoriek a následne do grafu vykreslite vzťah medzi počtom trénovacích vzoriek a úspešnosťou systému.

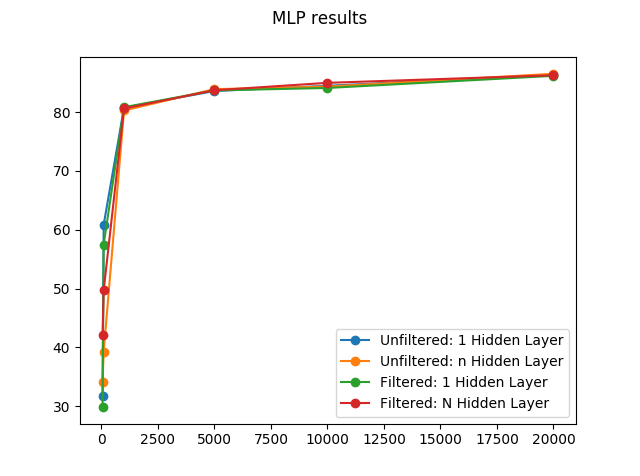
Pri MLP je potrebné vytvoriť architektúru s jednou skrytou vrstvou a n skrytými vrstvami (podľa vlastného výberu. Pri SVM je potrebné vyskúšať viacero kernelových funkcií(RBF, linear).

Porovnajte získané výsledky, kde ako vstup použijete dáta s duplikátmi a dáta bez duplikátov ( iba s 100, 1000, 20000 vzorkami ). Porovnanie úspešnosti medzi klasifikátormi, ktoré boli natrénované s duplikátmi a bez duplikátov vykreslite do grafu.

# Riešenie

## MLP

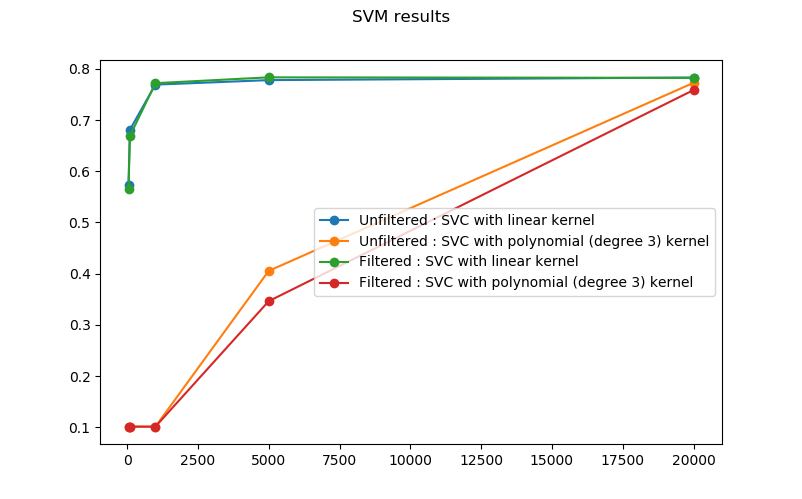
V tejto úlohe som naimplementoval strojové učenie pomocou algoritmu mlp z knižnice MLP.na natrénovanie dát som použil metódu mlp\_classifier(). Ako solver som sa rozhodoval medzi funkciami gsd a adam kvôli možnosti voľby viacerých parametrov oproti lbfg. Nakoniec som zvolil sgd a to z dôvodou, že adam požaduje veľke množstvá dát na dobré natr=novanie. Pomocou funmcie mlp\_predict do ktorej som vkladal testovacie dáta vzhodnotil % presnosť správneho odhladu vzhodnotených dát. Dáta som trénoval s jedným a dvoma neurónmi a to na všetých trénovacích dátach a aj filtrovaných trénovacích dátach, v ktorých su zmazané podobné obrázky.



Ako je možné vidieť z grafu, už pri poskytnutí 1000 trénovacích dát sme sa dostali na viac ako 80% úspešnosť vzhodnotenia dát a pri 20000 dátach na cca 95% úspešnosť.

## SVM

SVM som učil na dvoch rôznych kerneloch a to linerárnom a polynomiálnom. Úspešnosť rozpoznávania bola vyhodnocovaná na 10000 dátach a je možné ju vidieť nižšie na grafe:



Na grafe je možné vidieť, že lineárne SVM sa natrénovalo na menšom počte dát lepšie ako MLP ale ostalo približne na 80%. Polynomiálne SVM vyžaduje veľké množstvo dát na správne natrénovanie.