



باسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

مقدمه‌ای بر یادگیری ماشین – دکتر سید جمال‌الدین گلستانی

بهراد منیری – ۹۵۱۰۹۵۶۴

## گزارش تمرین سری ششم

## ۱ مقدمه

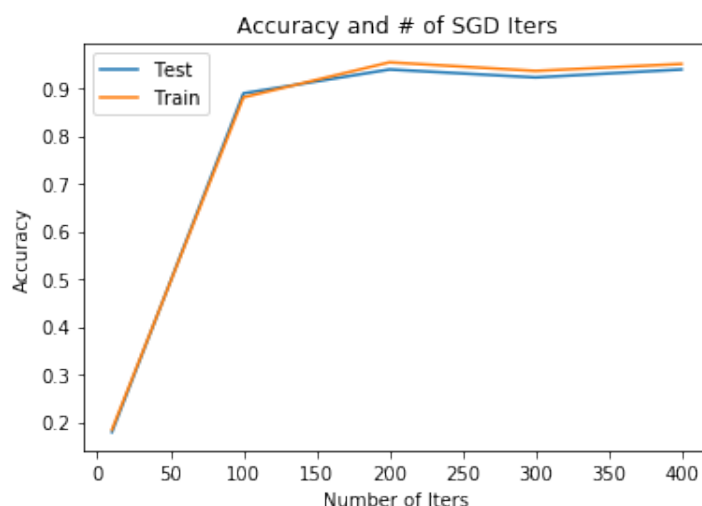
در این تمرین، از زبان پایتون برای پاسخ به سوالات استفاده شده و عمده‌ی توابع مورد استفاده از کتابخانه‌ی sklearn هستند.

در ابتدا، به کمک تابع `train_test_split()` داده‌های خود را به دو بخش آموزش (۹۰ درصد) و تست (۱۰ درصد) تقسیم می‌کنیم. در تمام طول این تمرین، داده‌های تست توسط الگوریتم یادگیری ماشین دیده نمی‌شوند و تنها برای ارزیابی فرضیه‌ی منتخب توسط الگوریتم یادگیری استفاده می‌شود.

ما از Classifier MLP پیاده‌شده در کتابخانه‌ی مذکور استفاده خواهیم کرد. الگوریتم یادگیری را Stochastic Gradient Descent قرار می‌دهیم با نرخ یادگیری‌ای برابر ثابت و برابر مقدار پیش فرض. همچنین تابع Activation را برابر ReLU قرار داده‌ایم.

## ۲ بررسی تاثیر تعداد قدم

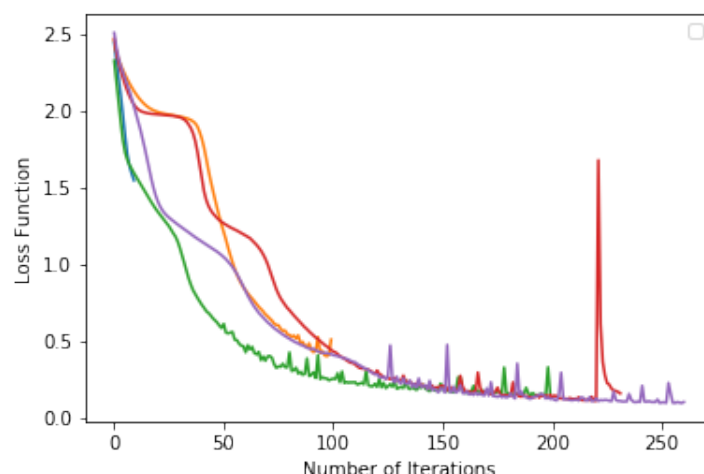
در این بخش، هشت لایه‌ی مخفی، هر کدام با هشت نورون، در نظر می‌گیریم. الگوریتم یادگیری را ۵ بار با تعداد قدم‌های {10, 100, 200, 300, 400} اجرا می‌کنیم.



شکل ۱: اثر تغییرات تعداد استپ‌های گرادینت

مشاهده می‌شود که با افزایش تعداد گام‌ها، درصد صحت بر داده‌های تست و ترین افزایش می‌یابد. در شکل زیر، روند تغییر تابع Loss را در هر یک از ران‌های فوق مشاهده می‌کنید.

همان‌طور که انتظار داریم، افزایش تعداد قدم باعث شده است که الگوریتم نقطه‌ی بهینه‌تری را پیدا کند که این موضوع خود

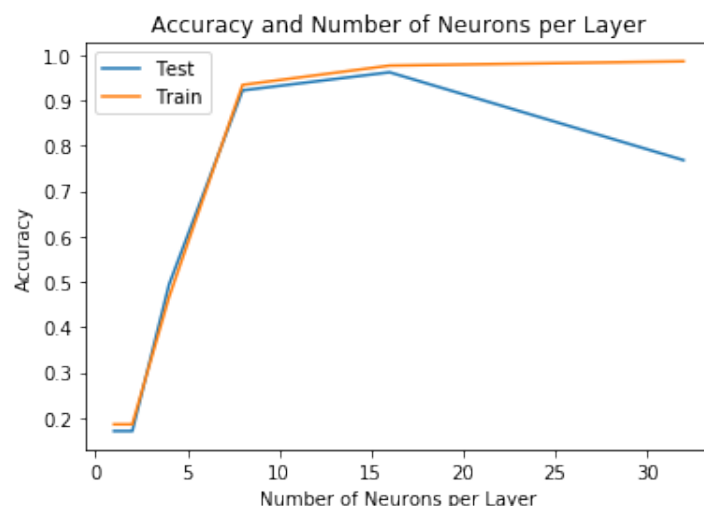


شکل ۲: مقدار تابع تلف بر حسب تعداد گام در پنج ران مختلف

را در خطای تجربی و خطای تست خود را نشان داده است. با ای مشاهدات، از آن جایی که از ۱۰۰ استپ به بعد، عملکرد بهبود چشم‌گیری نداشته است، بهتر است برای کاهش محاسبات، از همان ۱۰۰ (یا ۱۵۰) استپ گرادینان استفاده کنیم.

### ۳ بررسی تاثیر عرض شبکه

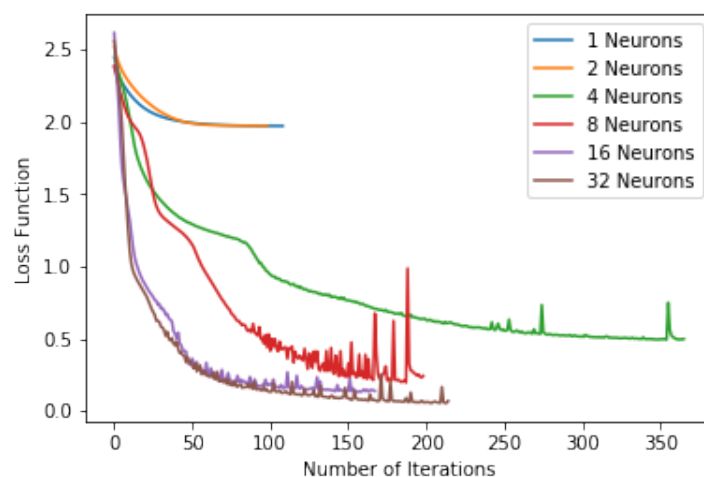
در این سوال به بررسی رابطه‌ی عمق شبکه و عملکرد آن بر روی داده‌های تست می‌پردازیم. شبکه‌ای با هشت لایه را با تعداد نوروں‌های هر لایه‌ی  $\{1, 2, 4, 8, 16, 32\}$  آموزش می‌دهیم. تعداد گام‌های الگوریتم SGD را برابر ۱۰۰ می‌گذاریم. عملکرد این طبقه‌بند را بر داده‌های تست و ترین، بر حسب تعداد نوروں‌های لایه‌ی مخفی، در شکل زیر مشاهده می‌کنید.



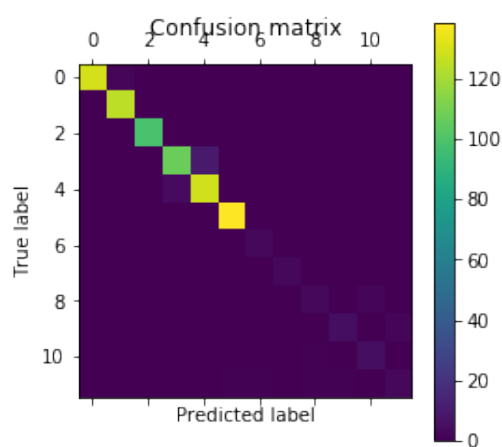
شکل ۳: عملکرد بر حسب تعداد نوروں‌های هر لایه در پنج ران مختلف

نمودار مقدار Loss بر حسب تعداد گام نیز در شکل زیر آمده است. انتظار داشتیم که با افزایش تعداد نوروں‌های هر لایه، عملکرد روی داده‌های ترین بهتر و بهتر شود که این اتفاق رخ داد زیرا در آن فضایی که تابع لاس را روی آن کمینه می‌کنیم بزرگتر شده است. همچنین انتظار داشتیم بعد از حدی، دچار مشکل overfitting شویم که در ۳۰ نوروں مخفی به وضوح این موضوع خود را در اختلاف صحت روی داده‌ی تست و ترین نشان داده است. بهترین عملکرد در ۱۶ نوروں مخفی رخ داده است.

در این بخش، ماتریس Confusion نیز رسم شده است تا درک بهتری از عملکرد طبقه‌بند داشته باشیم.



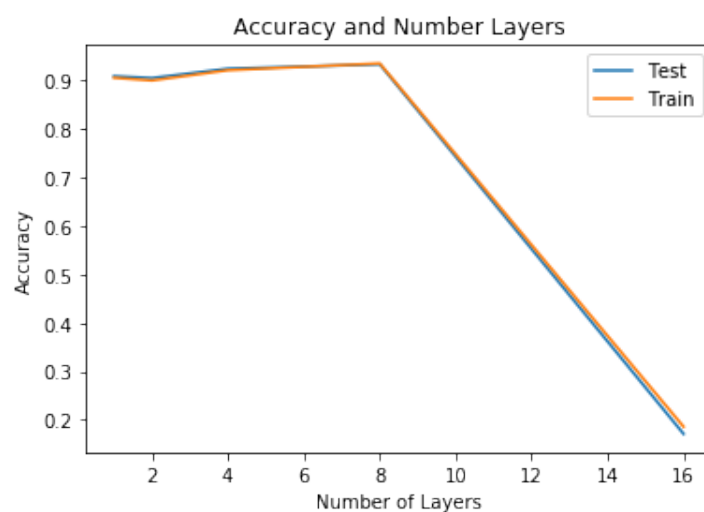
شکل ۴: مقدار تابع تلف بر حسب تعداد نورون‌های هر لایه در پنج ران مختلف



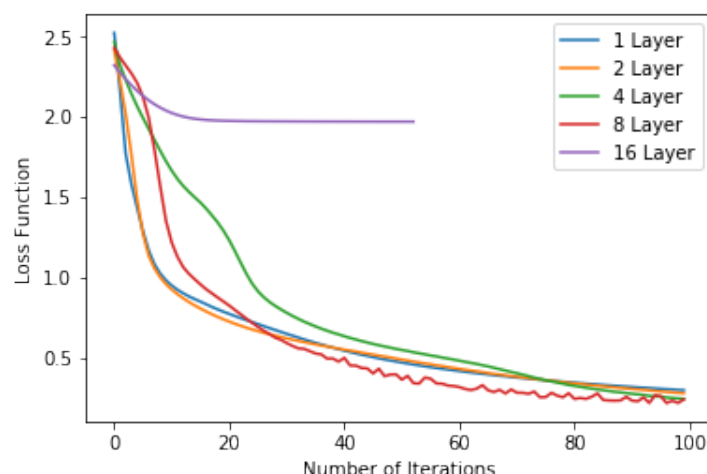
شکل ۵: ماتریس کانفیوژن

## ۴ بررسی تاثیر عمق شبکه

در این بخش، تعداد لایه‌ها را افزایش می‌دهیم و نتایج بخش قبل را مجدداً در این حالت نیز محاسبه می‌کنیم.



شکل ۶: عملکرد بر حسب تعداد لایه



شکل ۷: تلف بر حسب تعداد لایه

در شکل ۷ دیده می‌شود که در حالت ۱۶ لایه، الگوریتم SGD زود قطع می‌شود. دلیل این امر این است که مقدار تابع تلف آن بعد از چندین قدم، بهبود نمی‌یابد.

بیشینه دقت در هشت لایه رخ می‌دهد و بعد از آن دقت (حتی بر روی داده‌ی ترین) کاهش می‌یابد. دلیل این کاهش دقت این است که الگوریتم SGD برای یافتن نقطه‌ی بهینه در فضای بزرگتر نیاز به استپ‌های به مراتب بیشتری دارد و بهتر از الگوریتم‌های بهینه‌سازی‌ای مانند Adam استفاده کنیم. همان‌طور که مشاهده می‌شود در حدود ۸ لایه، بهترین عملکرد را داریم.