



۱. (۱۵٪) [پژوهش - شبکه مولد مقابله‌ای] در این پژوهش مروری بر کاربردهای شبکه مولد مقابله‌ای (GAN) در پردازش زبان طبیعی (حداقل دو مورد) انجام دهید و نحوه استفاده از این شبکه در آن کاربردها را بیان کنید.

۲. (۱۵٪) [پژوهش - شبکه عصبی مبدل] در این سوال، علاوه بر GPT-3، مروری بر حداقل دو مدل زبانی بسیار بزرگ آموزش داده شده مبتنی بر مبدل (Transformer) مانند Megatron-Turing داشته باشید و نمونه‌هایی از کاربردهای آنها را بیان کنید.

۳. (۷۰٪) [پیاده‌سازی - برچسب‌زنی اجزای کلام با LSTM] از شبکه عصبی LSTM برای برچسب‌زنی اجزای کلام فارسی استفاده کنید. برای این کار از دادگان پیوست شده به این تمرین استفاده کنید. در این دادگان ۵۰۰ هزار واحدی، ۵ هزار واحد را به صورت تصادفی (انتخاب جملات تصادفی) برای آزمون جدا کنید و مابقی را برای آموزش به کار بگیرید. برای ورودی شبکه از بردار کلمات استفاده کنید و این بردارها را با روش Word Embedding که برای زبان فارسی روی اینترنت هست (مانند Word2Vect, FastText, Bert) استفاده کنید. در پیاده‌سازی این سوال می‌توانید از کتابخانه‌ها و ابزارهای آماده استفاده کنید.

الف) یک شبکه یک طرفه با یک لایه مخفی و تعداد مناسب سلول حافظه در لایه مخفی (مثلاً ۱۵۰) بسازید. مقدار درستی (Accuracy) الگوریتم را برای مجموعه آموزش و آزمون بدست آورید. نمودار Loss را برای تکرارهای مختلف در حین آموزش رسم کنید.

ب) از یک شبکه دوطرفه با تعداد مناسب سلول حافظه (مثلاً ۸۰) استفاده کنید و مقدار درستی را روی داده آموزش و آزمون بدست آورده و با نتایج قسمت الف مقایسه کنید.

ج) در شبکه قسمت ب، علاوه بر بردار کلمات، از بردار One-hot برچسب‌ها برای برچسب کلمه قبلی نیز به عنوان یکی دیگر از ورودی‌ها استفاده کنید و کارایی روش را در این حالت نیز برای داده‌های آموزش و آزمون بدست آورید و نتایج را با قسمت ب مقایسه کنید.