

dimensionality
reduction استادمی

① ① برای حل مشکل نندین ابعاد معمولاً از روش‌های کاهش بعد یا

دوره بسیار ساده‌تر و بدلی کاهش بعد PCA و Autoencoder هسته.

PCA کاهش بعد را با یاد گرفتن یک تکب خطی (که داده را به فضای کوچکتری map می‌کند) یا دی‌گه

با محدود کردن بعد n به تعداد کمترین که بتواند واریانس دیتاست را بگیرد می‌توانیم به حذف کاهش بعد برسیم.

① - Autoencoder ها نیز با قراردادن چند لایه مخفی ابعاد را به یک فضای latent با تعداد بعد کمی بریزند

و چون تعداد بعد و Autoencoder کم است معمولاً داریم Data reduction انجام می‌دهیم.

① ② Word2Vec با آموزش یک شبکه عصبی یک بازتابی از داده یادگیری درونی در Glove
از روی متن (corpus) یک ماتریس ~~تکب~~ ^{ماتریس} ~~تکب~~ ^{ماتریس} سازد (matrix factorization)

در واقع زیربافت اصلی Glove این است که روی ~~تکب~~ ^{ماتریس} دیتاست دقت بسیار بالایی دارد در حال

که Word2Vec روی بقیه دیتاست‌ها بهتر عمل می‌کند.

به بیان دیگر Glove بهتر است چون ①

② به لغات خوب‌تر و فرکانس‌های کمتر در داده‌ها
دیده‌دار embedding معنای بیشتری می‌بخشد چون روی لغات

③ کمالات و در نظر می‌گیرد. ③ حذف آن خطی سراسر است (مجبور کردن به وارد کردن لغات که در فضای

روابط خطی دیده‌دار پس دقت‌های Glove بهتر عمل می‌کند

در حالی که Word2Vec روابط خطی را نمی‌تواند به خوبی تشخیص دهد.

Attention model

(۲) (۱)

نویس اصلی ساختار transformer این است که در هر مرحله با دسترسی مستقیم به تمام سطر قبلی داریم (self attention)

که علاوه بر این برای دود اطلاعات باقی می ماند علاوه بر آن سعی کنیم به اعلان های گذشته را نیز

در یک زمان نگاه کنیم که مدل ترنس bidirectional RNN را بدون (محاسبه آن) به همراه می آوریم.

روش کار: (مشخص می)

① Non-Sequential

② self-attention

attention در واقع یک گامینیم deep است که نتیجه ای از داده ها را مشخص می کند

(۲) (۲)

بعبود selective به معنی از آن ها توجه می کند مانند مثال ما در اسلاید است

در طای که self-attention به صورتی از داده های ورودی یک وزن می دهد که آن وزن را

مشاهدات قبل است و توجه بیشتر را به آن می دهد.