تمرین کوتاه ۸ شبکه های عصبی – مازندرانیان – ۸۳۰۴۰۲۰۶۶

سوال ۱) LDA چیست؟

LDA (مخفف Linear Discriminant Analysis) یک روش آماری و الگوریتم یادگیری ماشین است که در دسته‌بندی داده‌ها و کاهش بُعد استفاده می‌شود. این روش، به‌ویژه در مسائل طبقه‌بندی چندکلاسه محبوب است.هدف LDA پیدا کردن فضایی است که داده‌های متعلق به کلاس‌های مختلف را به بهترین شکل ممکن از یکدیگر جدا کند.

کاربرد ها:

- **طبقه‌بندی (Classification):** استفاده به عنوان الگوریتم طبقه‌بندی برای پیش‌بینی کلاس‌های جدید.
- **کاهش بُعد (Dimensionality Reduction):** کاهش تعداد ویژگی‌ها در داده‌های با ابعاد بالا، در حالی که اطلاعات مربوط به کلاس‌ها حفظ شود.
- **پیش‌پردازش داده‌ها :** اغلب برای کاهش بُعد قبل از استفاده از الگوریتم‌های یادگیری ماشین مانند SVM یا Logistic Regression به کار می‌رود.

سوال ۲) اثبات فرمول زیر:

12 of 29

$$L(v_1; Z) = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i - (x_i^\top v_1)v_1\right\|_2^2$$

- Can be shown to be equivalent to maximizing variance:

$$L(v_1; Z) = -\mathrm{Var}\left(\{x_i^\top v_1\}_i\right)$$

- If variance of projection on v_1 is low, v_1 is not informative about x_i

مرحله اول: تعریف و تجزیه $L(v_1; Z)$

معادله $L(v_1; Z)$ به صورت زیر تعریف شده است:

$$L(v_1; Z) = \frac{1}{n} \sum_{i=1}^{n} \|x_i - (x_i^\top v_1)v_1\|_2^2$$

اینجا:

$(x_i^\top v_1)v_1$ تصویر x_i روی v_1 است.

$x_i - (x_i^\top v_1)v_1$ بردار باقی‌مانده (Residual) است.

مرحله دوم: رابطه بین x_i و تصویر آن

مربع نُرم تفاضل به صورت زیر محاسبه می‌شود:

$$\|x_i - (x_i^\top v_1)v_1\|_2^2 = \|x_i\|_2^2 - \|(x_i^\top v_1)v_1\|_2^2$$

از آنجایی که v_1 یک بردار واحد است ($\|v_1\|_2 = 1$)، داریم:

$$\|(x_i^\top v_1)v_1\|_2^2 = (x_i^\top v_1)^2$$

بنابراین:

$$\|x_i - (x_i^\top v_1)v_1\|_2^2 = \|x_i\|_2^2 - (x_i^\top v_1)^2$$

مرحله سوم: محاسبه کل $L(v_1; Z)$

با جایگذاری این مقدار در معادله اصلی داریم:

$$L(v_1; Z) = \frac{1}{n}\sum_{i=1}^{n}\left(\|x_i\|_2^2 - (x_i^\top v_1)^2\right)$$

این عبارت را می‌توان به دو بخش تفکیک کرد:

$$L(v_1; Z) = \frac{1}{n}\sum_{i=1}^{n}\|x_i\|_2^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i^\top v_1)^2$$

بخش اول ($\frac{1}{n}\sum_{i=1}^{n}\|x_i\|_2^2$) یک مقدار ثابت است و مستقل از v_1 است. بنابراین، تنها بخش دوم مهم است:

$$L(v_1; Z) \propto -\frac{1}{n}\sum_{i=1}^{n}(x_i^\top v_1)^2$$

مرحله چهارم: ارتباط با واریانس

واریانس یک مجموعه داده تعریف می‌شود به صورت:

$$\mathrm{Var}\left(\{x_i^\top v_1\}_i\right) = \frac{1}{n}\sum_{i=1}^{n}(x_i^\top v_1)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}x_i^\top v_1\right)^2$$

چون v_1 را به‌گونه‌ای انتخاب می‌کنیم که $x_i^\top v_1$ دارای میانگین صفر باشد (مرکز داده‌ها در PCA)، مقدار میانگین صفر می‌شود:

$$\frac{1}{n}\sum_{i=1}^{n}x_i^\top v_1 = 0$$

در نتیجه:

$$\mathrm{Var}\left(\{x_i^\top v_1\}_i\right) = \frac{1}{n}\sum_{i=1}^{n}(x_i^\top v_1)^2$$

بنابراین:

$$L(v_1; Z) \propto -\mathrm{Var}\left(\{x_i^\top v_1\}_i\right)$$

نتیجه‌گیری

اثبات کردیم که:

$$\mathrm{Var}\left(\{x_i^\top v_1\}_i\right) - = L(v_1; Z)$$

اگر نیاز به جزئیات بیشتر یا توضیح یک مرحله خاص دارید، اطلاع دهید!

سوال ۳) اثبات فرمول زیر:

1D Case

- The principal components analysis (PCA) algorithm computes

$$v_1^* = \min_{v_1} L(v_1; Z) = \max_{v_1} v_1^\top C v_1$$

- **Theorem:** Solution is $v_1^* = \mathrm{TopEigenvector}(C)$
  - That is, eigenvector corresponding to the largest eigenvalue
  - **Recall:** If $Cv = \lambda v$, then v is an eigenvector corresponding to eigenvalue λ

اثبات اینکه v_1^* بزرگ‌ترین بردار ویژه C است

1. فرموله کردن مسئله به زبان ماتریسی

ماتریس کواریانس C به صورت زیر تعریف می‌شود:

$$C = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^\top$$

که x_i داده‌ها هستند.

هدف PCA این است که بردار v_1 را پیدا کند که واریانس $x_i^\top v_1$ را بیشینه کند:

$$\mathrm{Var}(\{x_i^\top v_1\}) = v_1^\top C v_1$$

2. شرط واحد بودن v_1

برای اینکه v_1 یک بردار معتبر باشد، شرط $\|v_1\|_2 = 1$ را اضافه می‌کنیم. بنابراین مسئله به‌صورت زیر بازنویسی می‌شود:

$$v_1^* = \arg\max_{v_1} v_1^\top C v_1 \quad \text{subject to } \|v_1\|_2 = 1.$$

3. استفاده از ضرب‌لگاریج (Lagrange Multipliers)

برای حل این مسئله با شرط محدودیت $\|v_1\|_2 = 1$، از روش ضرب‌لگاریج استفاده می‌کنیم. تابع لگاریج به‌صورت زیر تعریف می‌شود:

$$\mathcal{L}(v_1, \lambda) = v_1^\top C v_1 - \lambda(v_1^\top v_1 - 1)$$

که λ ضریب لگاریج است.

4. مشتق‌گیری از $\mathcal{L}$

برای یافتن مقدار بهینه v_1^*، مشتق تابع لگاریج را نسبت به v_1 و λ برابر صفر قرار می‌دهیم:

- مشتق نسبت به v_1:

$$\frac{\partial \mathcal{L}}{\partial v_1} = 2Cv_1 - 2\lambda v_1 = 0$$

که معادل است با:

$$Cv_1 = \lambda v_1$$

- مشتق نسبت به λ:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(v_1^\top v_1 - 1) = 0$$

که معادل است با:

$$v_1^\top v_1 = 1$$

5. نتیجه‌گیری از معادلات

از $Cv_1 = \lambda v_1$ نتیجه می‌گیریم که v_1 یک بردار ویژه ماتریس C است و λ مقدار ویژه متناظر با آن است.

6. انتخاب بزرگ‌ترین مقدار ویژه

از آنجا که هدف PCA بیشینه کردن $v_1^\top C v_1$ است، λ باید بزرگ‌ترین مقدار ویژه C باشد. بنابراین، v_1^* بردار ویژه متناظر با بزرگ‌ترین مقدار ویژه $\lambda_{\max}$ خواهد بود.