درس شبکه های عصبی دکتر مسعودنیا - تمرین 1 - مازندرانیان - 830402066

سوال 1) دلیل معکوس ناپذیری ماتریس $x^T x$

به دو دلیل:

1) چند خطی بودن داده ها

به عنوان مثال در دیتاست دو ستون وزن و سن داریم، رابطه ی بین این دو به این شکل است که هر چه سن افراد بیشتر باشد وزن بیشتری هم دارند، ماتریس زیر رو در نظر بگیرید:

X = | 1 2 | | 2 4|

مثلا در این ماتریس قد دقیقا دو برابر سن است، اگر $x^T x$ رو حساب کنیم:

X^TX = | 5 10 | | 10 20|

همانطور که میبینید دترمینان این ماتریس صفر است پس معکوس پذیر نیست. توی این حالت به اصطلاح میگیم ویژگی سن و وزن با هم همبستگی دارند.
برای رفع این مشکل چند راه داریم: حذف یکی ویژگی ها، استفاده از رگرسیون ridge و یا loss، جمع آوری داده های بیشتر و یا دستکاری داده های موجود

2) سیستم نامعین

به حالتی گفته می شود که تعداد ویژگی های ما بیشتر از تعداد داده ها باشد، یا به عبارتی تعداد مجهولات معادله از تعداد آن معادله بیشتر است، مثلا:

$x + y = 5$

ما با داشتن تنها یک معادله ی بالا نمی توانیم برای مقادیر x و y یک مقدار واحد پیدا کنیم. برای حل این سیستم می توانیم یکی از متغیر ها را به عنوان متغیر آزاد انتخاب کنیم یعنی $x = 5 - y$ به این شکل برای مقادیر مختلف y میتوانیم مقدار x رو بدست بیاریم.

سوال 2) روش های تخمین ماتریس $x^T x$

معمولا مقدار این ماتریس را تخمین میزنیم به دلایلی مثل افزایش سرعت و کارایی، ماتریس های نامعین و در ماتریس های بزرگ که ممکن از در محاسبات دچار خطای گرد کردن بشیم.
از روش های عددی رایج میتونیم به این موارد اشاره کنیم:

1) تجزیه LU: ماتریس $x^T x$ به حاصل ضرب دو ماتریس مثلثی پایین و بالا تجزیه می‌شود. سپس معکوس هر یک از این ماتریس‌ها به راحتی محاسبه می‌شود و در نهایت معکوس ماتریس اصلی به دست می‌آید.
2) تجزیه QR: ماتریس $x^T x$ به حاصل ضرب یک ماتریس متعامد و یک ماتریس مثلثی فوقانی تجزیه می‌شود. این روش برای ماتریس‌های بزرگ و نامعین بسیار مفید است.
3) روش های تکرار شونده: ین روش‌ها با یک تقریب اولیه شروع می‌کنند و با تکرار یک سری عملیات، تقریب را بهبود می‌بخشند. روش‌های تکرارشونده مانند روش گرادیان مزدوج و روش گاوس-سایدل از جمله روش‌های محبوب در این زمینه هستند.
4) روش های مبتنی بر svd: با استفاده از تجزیه SVD ماتریس $x^T x$، می‌توان معکوس کاذب مور-پنروز را محاسبه کرد که در مواردی که ماتریس نامعین است، بسیار مفید است.

نکته 1) تفاوت گرادیان مزدوج یا conjugate با گرادیان نزولی یا descent این است که گرادیان مزدوج در هر مرحله شیب جدید رو عمود بر جهت شیب قبلی انتخاب می شود که باعث میشود سریع به همگرایی برسیم.
نکته 2) یکی از راه های مشکل چند خطی بودن دیتا این است که از ridge regression برای پیدا کردن ضرایب استفاده کنیم.
در این روش با تعریف یک متغیر λ به عنوان یک وزن برای محاسبه میزان خطا استفاده میکنیم. هر چه λ بزرگتر باشد ضرایب به سمت صفر سوق داده میشود و هر چه کوچکتر باشد جریمه کوچکتر میشود و مدل ما به رگرسیون خطی معمولی نزدیکتر میشود.
برای پیدا کردن مقدار مناسب λ میتونیم از روش های cross validation و grid search و استفاده کنیم.

نکته 3) یکی از راه های رفع undetermined system استفاده از شبه معکوس ماتریس است.

A = | 1 2 | | 2 4|

A^+ = | 0.4 -0.2 | | -0.2 0.1|

خواص معکوس شبه:

- اگر ماتریس A مربعی و معکوس‌پذیر باشد، معکوس شبه آن با معکوس معمولی آن برابر است.
- شبه معکوس یک ماتریس همیشه وجود دارد و یکتا است.
- شبه معکوس یک ماتریس متعامد نیست، مگر در صورتی که ماتریس اصلی متعامد باشد.

$$\text{Linear Regression}: \sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2$$

$$\text{Lasso Regression}: \sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

$$\text{Ridge Regression}: \sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}{\beta_j}^2$$