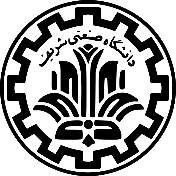
به نام خدا

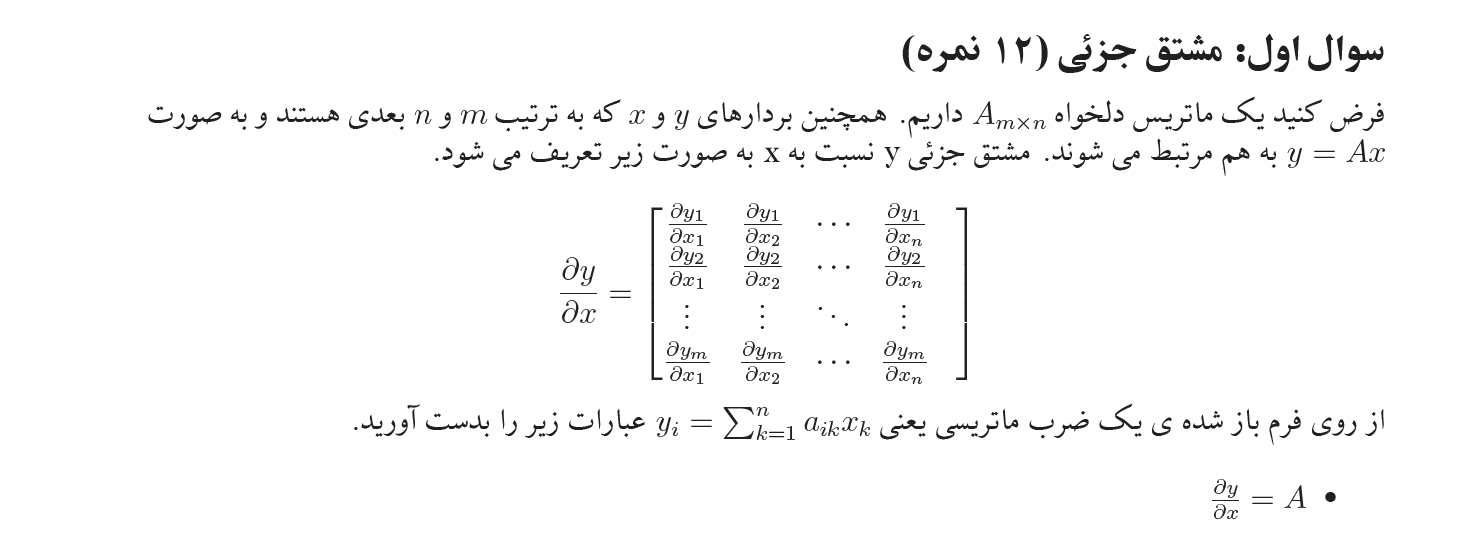


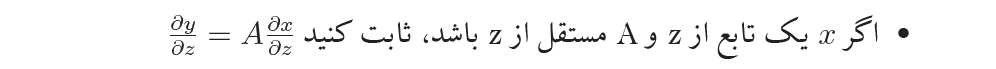
**یادگیری ژرف**

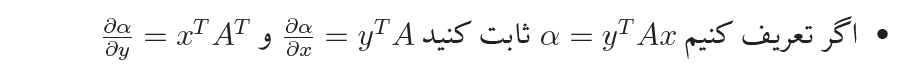
**ﻧﯿﻢﺳﺎﻝ دوم 03 -02**

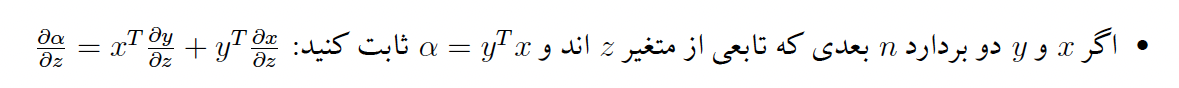
ﻣﺪﺭﺱ: دکتر مهدیه سلیمانی

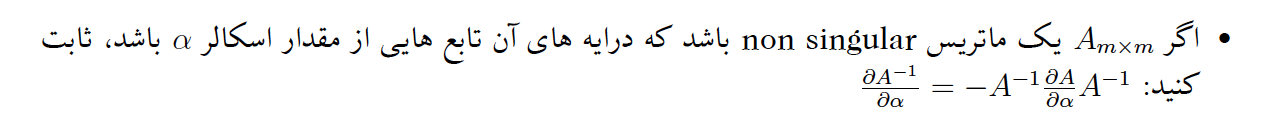
سید امیر کسائی- 40221221- همفکری با: امیر محمد عزتی

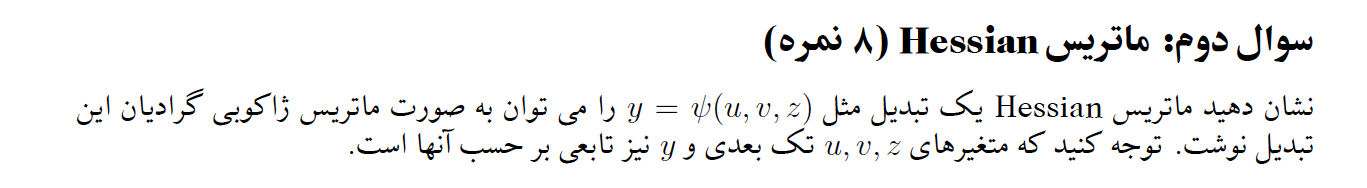


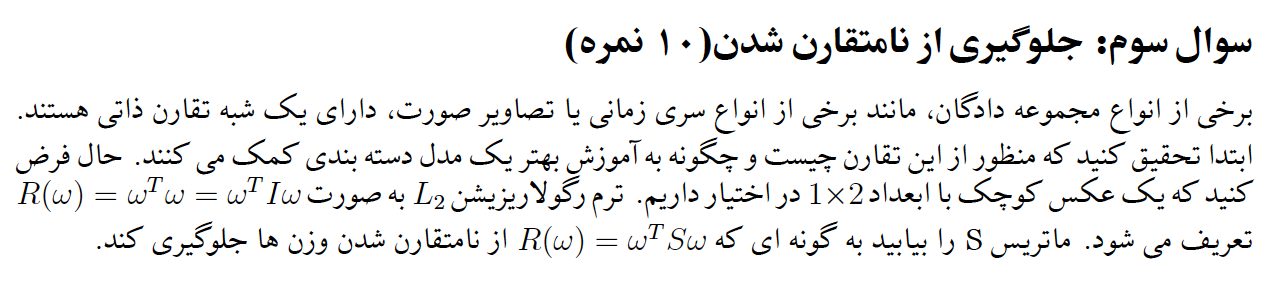
**











شبه تقارن ذاتی داده ها به معنای وجود الگو ها یا تقارن های ساختاری در خود داده هاست که می‌تواند برای درک و مدل سازی بهتر مورد استفاده قرار گیرد. از مثال های آن در حوزه تصویر می‌توان به تقارن چرخشی،‌تقارن انعکاسی و تقارن translational اشاره کرد. در تقارن چرخشی جسم پس از درجه خاصی از چرخش مثل حالت اولیه به نظر می‌رسد. در تقارن translational الگو هایی مانند آنچه در کاشی ها است وجود دارد و در واقع شکل دارای واحد های تکرار شونده است. تقارن انعکاسی مانند تصاویر پروانه ها است که در امتداد محور خاصی تقارن وجود دارد. در حوزه متن هم می‌توان به وجود ساختار متقارن از لحاظ معنا یا وجود جفت های واژگان مانند مترادف و متضاد اشاره کرد. داده های سری زمانی هم می‌توانند دارای تقارن باشند. مانند داده های فصلی. شناسایی این تقارن ها کاربرد های مختلفی در زمینه مدل سازی دارد.

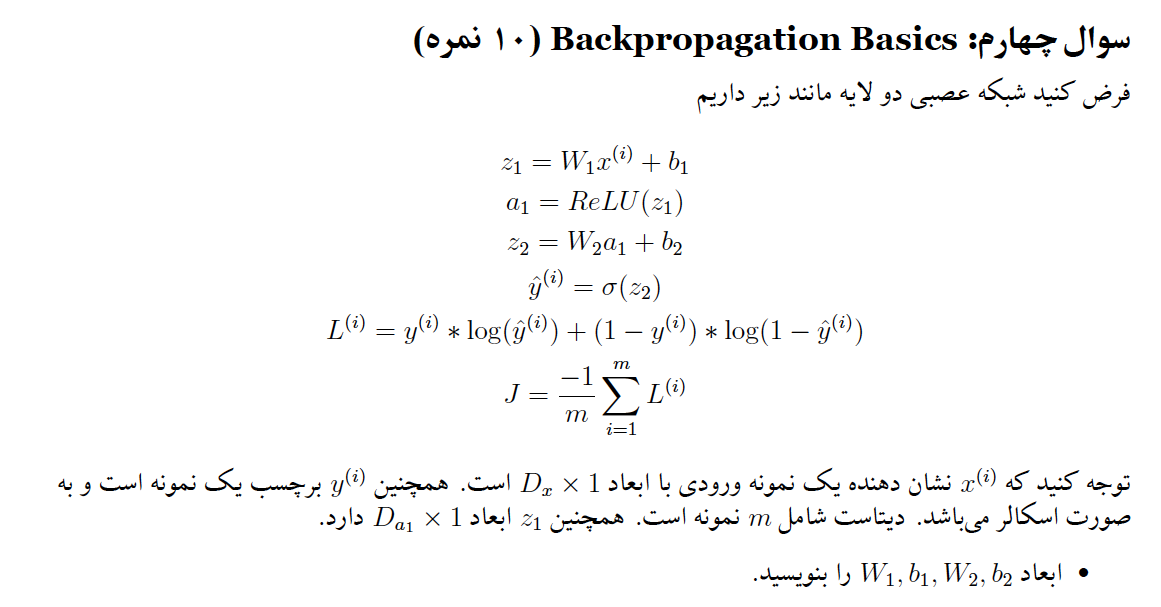
Regularization: درک شبه تقارن ها می‌توان منجر به استفاده بهتر از تکنیک های منظم سازی شود که وزن های شبکه را تشویق به یادگیری تقارن ها می‌کند.

Feature representation: شبه تقارن ذاتی می‌تواند در انتخاب یا حتی ایجاد ویژگی ها کمک کننده باشد. با شناسایی تقارن ها،‌محققان می‌توانند ویژگی هایی طراحی کنند که به یادگیری الگو های متمایز توسط مدل کمک کنند.

Data Augmentation: با اعمال چرخش یا بازتاب یا پیدا کردن الگو های تکرار شونده می‌توان داده های جدیدی تولید کرد.

معماری مدل: برای مثالی از این کاربرد می‌توان به شبکه های عصبی CNN اشاره کرد که از تقارن در تصاویر استفاده می‌کنند.

تفسیر پذیری: درک تقارن ذاتی در داده ها می‌تواند منجر به داشتن مدل های قابل تفسیر بیشتری شود. زیرا با شناسایی و مدل سازی این تقارن ها می‌توان بینش خوبی درباره ساختار داده ها و عوامل موثر در طبقه بندی بدست آورد.

**

**

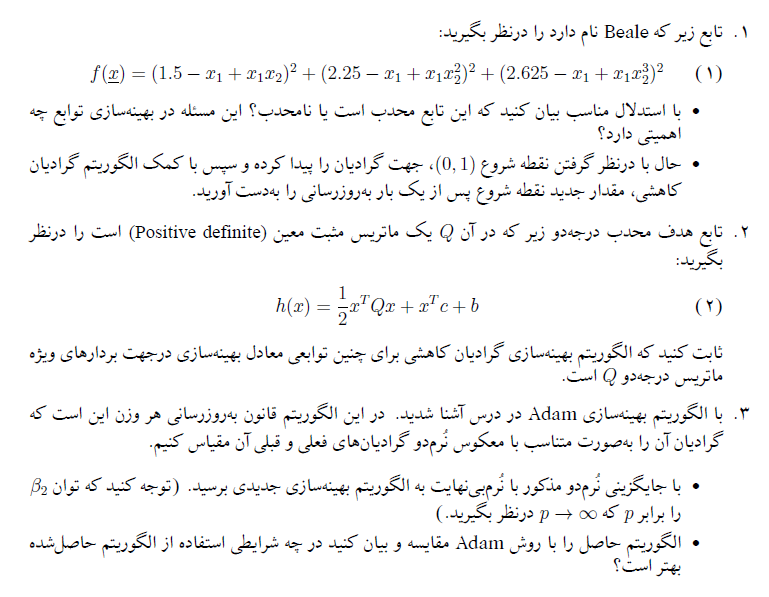
**

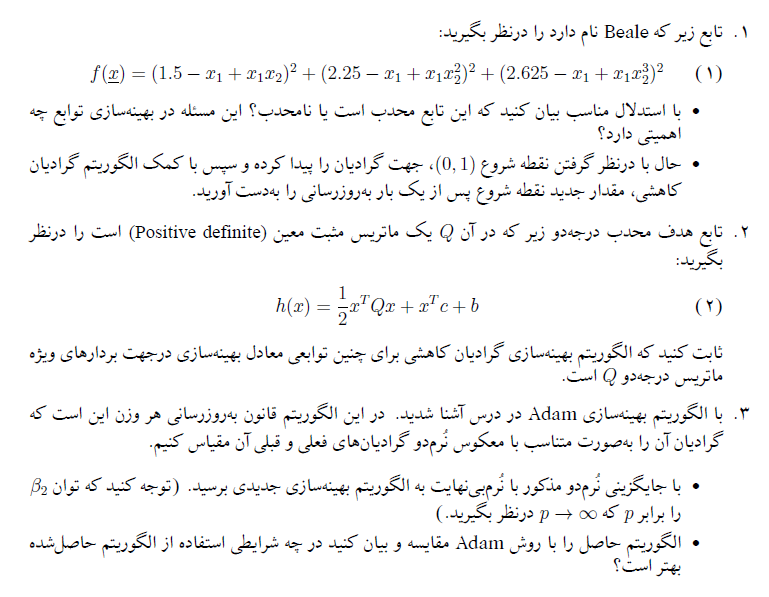
**

**

**

**

**سوال پنجم: بهینه سازی (۲۰ نمره)**



* 1. در مورد تابع Beale، یک saddle point در (0,1) دارد، که نقطه ای است که نه بیشینه محلی است و نه کمینه محلی.  
     saddle point مکانی است که در آن تابع در یک جهت منحنی رو به بالا می شود و در جهت دیگر منحنی رو به پایین می شود، مانند زین و این ویژگی توابع غیر محدب است.

1. برای سادگی فرض میکنیم :

برای بررسی توان در ماتریس بدست آمده، فرض میکنیم () یک eigenvalue decompositionبرای است که در آن ماتریسی متعامد () از بردارهای ویژه و یک ماتریس قطری از مقدار های ویژه است. بنابراین داریم:

از آنجایی که گرادیان کاهشی نسبت به جابجایی تغییر ناپذیر است پس این رابطه برای نیز صادق است.

بنابراین گرادیان کاهشی معدل بهینه سازی در جهت پردار های ویژه ماتریس Q (S) است.

[CS Toronto](https://www.cs.toronto.edu/~rgrosse/courses/csc421_2019/slides/lec07.pdf)

* 1. الگوریتم Adam به صورت زیر است:

(Get gradients w.r.t. stochastic objective at timestep )

(Update biased first moment estimate)

(Update biased second raw moment estimate(

(Compute bias-corrected first moment estimate)

(Compute bias-corrected second raw moment estimate)

(Update parameters)

در الگوریتم جدید (Adamax) داریم:

در نتیجه داریم:

(Update biased first moment estimate)

(Update the exponentially weighted infinity norm)

(Update parameters)

* 1. Adam و Adamax الگوریتم های بهینه سازی هستند که در آموزش مدل های یادگیری عمیق استفاده می شوند. Adam مزایای AdaGrad و RMSProp را ترکیب می کند و نرخ یادگیری را بر اساس لحظات اول و دوم گرادیان تنظیم می کند. Adamax، گونه‌ای از Adam بر اساس نورم بی‌نهایت، از حداکثر مقدار مطلق برای مقیاس‌بندی گرادیان‌ها استفاده می‌کند، و آن را به طور بالقوه پایدارتر و قوی‌تر می‌کند، به خصوص در موقعیت‌هایی با sparse gradients، noisy optimization، یا زمانی که گرفتن وابستگی‌های long-term بسیار مهم است. در حالی که Adam به طور گسترده برای کارایی خود در طیف گسترده ای از وظایف استفاده می شود، Adamax ممکن است به دلیل رویکرد متفاوت خود در مقیاس گرادیان، مزایایی را در سناریوهای خاص ارائه دهد.

[Adamax](https://arxiv.org/pdf/1412.6980v9.pdf)