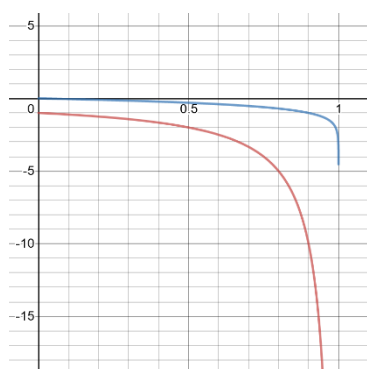


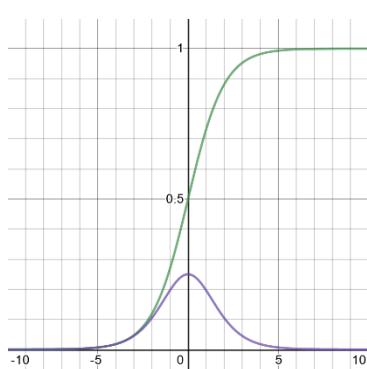
اصلاحیه :

توضیحات داده شده در کلاس دستیار آموزشی در مورد دلیل vanishing gradient در تابع خطای generator مقداری گمراه کننده میباشد و توضیحات زیر را در این مورد بررسی کنید که دچار اشتباه در فهم مطالب نشوید.

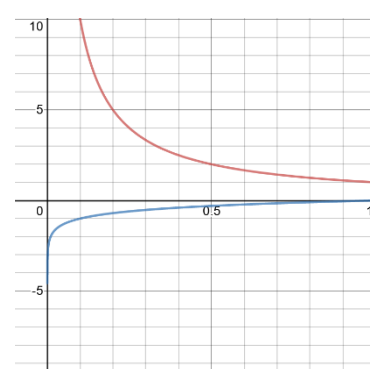
در تصویر اول توابع خطا و مشتق آن ($\log(1-x)$ و $\frac{\partial}{\partial x}(\log(1-x))$) کشیده شده است. که در این نمودارها خروجی Discriminator یا همان $D(x)$ می باشد. زمانی که Discriminator بتواند به خوبی داده های مربوط به generator را تشخیص دهد. مقدار خروجی آن نزدیک به صفر خواهد شد (و در تصویر دوم که خروجی سیگموئید می باشد (نمودار سبز رنگ) ، زمانی خروجی Discriminator نزدیک به صفر می باشد که خروجی سیگموئید نیز نزدیک به صفر می باشد) و چون مشتق سیگموئید نیز در آن نقاط نزدیک به صفر می باشد زمانی که شبکه در مرحله ی آپدیت شدن می باشد که با ضرب مشتق تابع خطا ($\frac{\partial}{\partial x}(\log(1-x))$) در مشتق سیگموئید ($\frac{\partial}{\partial x}(\sigma(x))$) گرادینت خطا را به آخرین لایه ی شبکه باز میگردانیم. چون مقدار مشتق خطا عددی نزدیک به منفی ۱ میباشد و مشتق سیگموئید نزدیک به صفر میباشد در نتیجه باعث میشود که مشتق به صفر نزدیک شود و مشکل vanishing gradient ایجاد شود. اما اگر تابع خطا و گرادینت آن را به صورت $\log(x)$ و $\frac{\partial}{\partial x} \log(x)$ داشته باشیم. در نتیجه چون زمانی که مقدار Discriminator نزدیک به صفر باشد مقدار گرادینت آن بزرگ می باشد در نتیجه زمانی که با گرادینت سیگموئید ضرب میشود (همانطور که گفته شد عدد کوچکی می باشد زمانی که discriminator نزدیک به صفر می باشد) حاصل ضرب عددی نزدیک به صفر نمیشود و مشکل vanishing gradient در آپدیت generator پیش نخواهد آمد.



$\log(1-x)$ $\frac{\partial}{\partial x}(\log(1-x))$
تصویر اول



$\sigma(x)$ $\frac{\partial}{\partial x}(\sigma(x))$
تصویر دوم



$\log(x)$ $\frac{\partial}{\partial x} \log(x)$
تصویر سوم

در مورد سوالی که یکی از دانشجویها پرسیدند که FID چگونه Quality و Divergence را اندازه گیری میکند :

این را نیز به پاسخم اضافه میکنم که همانطور که توضیح داده شد. ابتدا تصاویر اصلی و تصاویر تولید شده را با استفاده از یک شبکه به فضای feature میبریم و در آن جا هر کدام از مجموعه featureها را با یک مدل گاوسی تخمین میزنیم. و فاصله ی این دو مدل را به عنوان FID در نظر میگیریم.

اگر داده‌ها دارای Divergence کمی باشند در نتیجه feature‌های هر کدام از این داده‌ها در فضای feature نزدیک به هم خواهند بود و گاوسی مربوط به این feature‌ها از گاوسی مربوط به داده‌های اصلی از هم فاصله میگیرند. و اما اگر داده‌ها Quality خوبی نداشته باشند feature‌های مربوط به داده‌ها در مکانی دور از feature‌های داده‌های اصلی قرار میگیرند و در نتیجه باز هم فاصله‌ی دو مدل از هم زیاد خواهد شد. و FID همزمان هر دو مقدار Divergence و Quality را ارزیابی میکند. به این معنی که اگر FID عدد بزرگی باشد یعنی Quality یا Divergence و یا هر دوی آنها مطلوب نمی‌باشند و اگر مقدار FID کم باشد یعنی داده‌های تولید شده از Quality و Divergence بهتری برخوردار میباشند.