1. محاسبه احتمال :

قبل از بررسی سوال، در ابتدا باید یکسری از فرمول های مرتبط برای حل را بیان کرد:

* 
* 
* 

حال ما میدانیم که اگر به ازای ورودی X، کلاس ما برابر با 1 شده باشد این به این منظور است که :

P(X|1) > P(X|0)

پس کافیست این مفاهیم را برای 4 عبارت اولیه که به ما داده شده است، توسعه دهیم :

P(aa|0) > P(aa|1) → P(aa|0) = P(a|0)\* P(a|0) > P(a|1)\* P(a|1) => P(a|0) > P(a|1)

P(ab|0) > P(ab|1) → P(ab|0) = P(a|0)\* P(b|0) > P(a|1)\* P(b|1)

P(ba|0) > P(ba|1) → P(ba|0) = P(b|0)\* P(a|0) > P(b|1)\* P(a|1)

P(bb|0) < P(bb|1) → P(bb|0) = P(b|0)\* P(b|0) > P(b|1)\* P(b|1) => P(b|0) > P(b|1)

* P(1) = 0.75
* P(0) = 0.25
* P(a|0) = 5/8
* P(b|0) = 3/8
* P(a|1) = 1/4
* P(b|1) = ¾
* aabaa 0: (3/4)\*(5/8)\*(5/8)\*(3/8)\*(5/8)\*(5/8) = 0.042

1: (1/4)\*(1/4)\*(1/4)\*(3/4)\*(1/4) \*(1/4) = 0.0007

<= عبارت بالا مربوط به کلاس0 است

* b 0: (3/4)\*(3/8)= 0.28

1: (1/4)\*(3/4) = 0.18

<= عبارت بالا مربوط به کلاس0 است

* bba 0: (3/4)\*(3/8)\*(3/8)\*(5/8) = 0.065

1: (1/4)\*(3/4)\*(3/4)\*(1/4) = 0.035

<= عبارت بالا مربوط به کلاس0 است

* bbbb 0: (3/4)\*(3/8)\*(3/8)\*(3/8)\*(3/8) = 0.015

1: (1/4)\*(3/4)\*(3/4)\*(3/4)\*(3/4) = 0.08

<= عبارت بالا مربوط به کلاس1 است

1. توضیح در مورد هر دیتاست:

**MNIST**: پایگاه‌داده MNIST (پایگاه‌داده مؤسسه ملی استاندارد و فناوری اصلاح شده) یک پایگاه‌داده بزرگ از ارقام دست‌نویس است که معمولاً برای آموزش سیستم‌های مختلف پردازش تصویر استفاده می‌شود. پایگاه‌داده همچنین به طور گسترده برای آموزش و آزمایش در زمینه یادگیری ماشین استفاده می‌شود. با "آمیختن مجدد" نمونه‌های مجموعه‌داده‌های اصلی NIST ایجاد شد. سازندگان احساس کردند که ازآنجایی‌که مجموعه‌داده آموزشی NIST از کارمندان اداره سرشماری آمریکا گرفته شده است، درحالی‌که مجموعه‌داده آزمایشی از دانش‌آموزان دبیرستانی آمریکایی گرفته شده است، برای آزمایش‌های یادگیری ماشین مناسب نیست.

علاوه بر این، تصاویر سیاه‌وسفید از NIST نرمال‌سازی شدند تا در یک جعبه محدودکننده پیکسل ۲۸ در ۲۸ قرار بگیرند و ضد مستعار شوند که سطوح خاکستری را معرفی می‌کرد.

.پایگاه‌داده MNIST شامل 60000 تصویر آموزشی و ۱۰۰۰۰ تصویر آزمایشی است. نیمی از مجموعه آموزشی و نیمی از مجموعه تست از مجموعه‌داده آموزشی NIST گرفته شده است، درحالی‌که نیمی دیگر از مجموعه آموزشی و نیمی دیگر از مجموعه تست از مجموعه‌داده‌های آزمایشی NIST گرفته شده است.

**CIFAR-10**: مجموعه‌داده CIFAR-10 (مؤسسه تحقیقات پیشرفته کانادایی) مجموعه‌ای از تصاویر است که معمولاً برای آموزش الگوریتم‌های یادگیری ماشین و بینایی کامپیوتر استفاده می‌شود. این یکی از پرکاربردترین مجموعه‌داده‌ها برای تحقیقات یادگیری ماشینی است. مجموعه‌داده CIFAR-10 شامل ۶۰۰۰۰ تصویر رنگی ۳۲ در ۳۲ در ۱۰ کلاس مختلف است. ۱۰ کلاس مختلف نشان‌دهنده هواپیما، اتومبیل، پرنده، گربه، گوزن، سگ، قورباغه، اسب، کشتی و کامیون است. ۶۰۰۰ تصویر از هر کلاس وجود دارد.-

**FER-2013:** حاوی تقریباً 30000 تصویر RGB صورت با عبارات مختلف با اندازه محدود به 48×48 استو برچسب‌های اصلی آن را می‌توان به ۷ نوع تقسیم کرد:

عصبانی، انزجار، ترس، شاد، غمگین، سورپرایز، خنثی.

برچسب انزجار حداقل ۶۰۰ تعداد تصویر را دارد، درحالی‌که برچسب‌های دیگر هرکدام نزدیک به ۵۰۰۰ نمونه دارند.

1. to categorical:

در MNIST و CIFAR-10ما از to categorical برای تبدیل داده‌های آموزشی خود قبل از اینکه به مدل خود منتقل کنید، استفاده می‌کنید. اگر داده‌های آموزشی ما از رده‌ها به‌عنوان اعداد استفاده می‌کند، to categorical آن اعداد را در بردارهای مناسب برای استفاده با مدل‌ها تبدیل می‌کند و ما نمی‌توانید به‌سادگی یک مدل طبقه‌بندی را بدون آن آموزش دهید.

1. x\_train و y\_train:

در هر دیتاست ما نیازمند یک تصویر و برچسب آن هستیم. اصولا در x\_train مقدار عکس ها ذخیره میشوند و در y\_train مقدار برچسب ها قرار دارند.

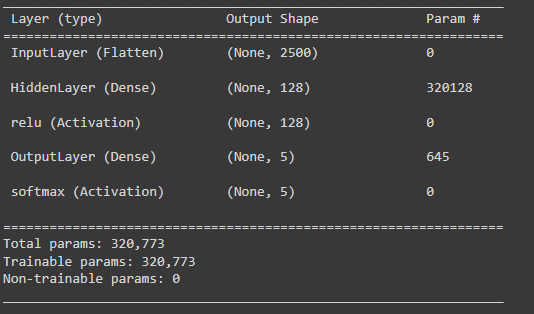
1. ImageDataGenerator:

این به ما امکان می دهد تا زمانی که مدل شما هنوز در حال آموزش است، تصاویر خود را در real-time افزایش دهید! ما می توانید هر گونه تبدیل تصادفی را روی هر تصویر آموزشی که به مدل منتقل می شود اعمال کنید. این نه تنها مدل شما را قوی می کند، بلکه در حافظه سربار نیز صرفه جویی می کند! و همین طور به ما امکان تولید دسته‌ها را می‌دهد: می‌توانید از تکه‌های کوچک‌تری از داده‌ها برای آموزش مدل خود با استفاده از نزول گرادیان دسته‌ای استفاده کنید. و عملاً همه مجموعه داده‌های دنیای واقعی مربوط به یادگیری عمیق معمولاً نمی‌توانند یکباره در حافظه جای بگیرند.

1. Summary ما نوع لایه ها ، تایپ آنهارا وهمینطور نام های مشخص میکند و علاوه بران موارد در وسط آن، مشخص میکند که چه shape را به عنوان خروجی تولید میکند. همینطور Param را نیز مشخص میکند که به طور مثال در حالتی که لایه ما HiddenLayer است از لایه قبلی 2500 ورودی به همراه bias دریافت میکند (2500+1) که چون در این لایه 128 تا unit داریم مقدار پارام مبا برای این لایه برابر با 2501\*128 میشود.

همینطور چون مقدار pre-defined شده ای نداریم، output های ما none هستند.

* Dense Layers:
  + output\_size \* (input\_size + 1) == number\_parameters
* For Conv Layers:
  + output\_channels \* (input\_channels \* window\_size + 1) == number\_parameters



1. در حالت کلی نباید انتظار داشت مدل به صورت خوبی عمل کند! برای این موضوع میتوان به دلایل زیر اشاره کرد :

* مدل ما تنها دارای یک لایه مخفی است و در این حالت که از دیتاست FER-2013 استفاده میکنیم، برعکس حالت MNIST، تصاویر ما دارای ویژگی های بیشتری هستند که آنهارا از یکدیگر متمایز میکنند و به همن دلیل نیاز داریم از تعداد لایه ها و تعداد نورون های بیشتری برای بهبود عملکرد بهره ببریم.
* تعداد epoch های اموزش بسیار پایین است.
* با این حال در حالت کلی برای اموزش یه NN برای این نوع تصاویر پیچیده که مدل ما باید چندین ویژگی را متمایز کنه، باید از CNN ها بهره ببریم و حتی با افزایش تعداد لایه های میانی و epoch به احتمال زیاد، عمکرد قابل قبولی نخواهند داشت.