

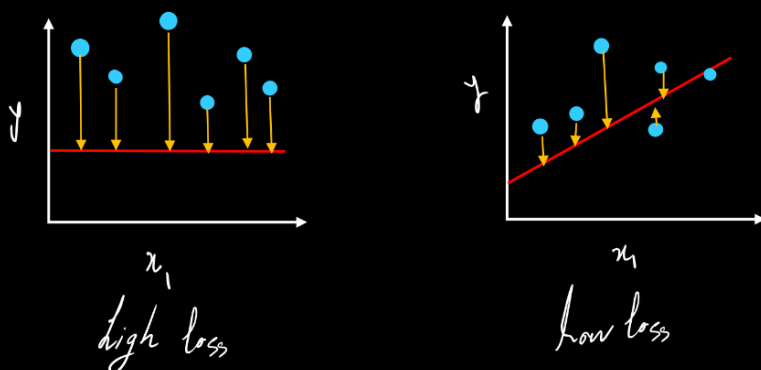
DESCENDING INTO ML: TRAINING AND LOSS

آموزش و تابع ضرر

آموزش (**Training**) یک مدل صرفاً به معنای یادگیری (تعیین) مقادیر خوب برای همه وزن‌ها و bias از نمونه‌های برچسب‌گذاری (labeled) شده است. در یادگیری نظارت شده (supervised learning)، یک الگوریتم یادگیری ماشینی با بررسی مثال‌های فراوان و تلاش برای یافتن مدلی که ضرر را به حداقل می‌رساند، یک مدل می‌سازد. این فرآیند به حداقل رساندن ریسک تجربی (**empirical risk minimization**) نامیده می‌شود.

ضرر مجازات یک پیش‌بینی بد است. یعنی ضرر عددی است که نشان می‌دهد پیش‌بینی مدل در یک مثال چقدر بد بوده است. اگر پیش‌بینی مدل کامل باشد، ضرر صفر است. در غیر این صورت ضرر بیشتر است. هدف از آموزش یک مدل، یافتن مجموعه‌ای از وزن‌ها و سوگیری‌هایی است که به طور متوسط در همه نمونه‌ها از دست دادن کم دارند. به عنوان مثال، به شکل زیر توجه کنید یک مدل با ضرر زیاد در سمت چپ و یک مدل با ضرر کم در سمت راست را نشان می‌دهد. در مورد شکل به موارد زیر توجه کنید:

- فلش‌ها نشان دهنده ضرر هستند.
- خطوط قرمز نشان دهنده پیش‌بینی‌ها هستند.



توجه داشته باشید که فلش‌های موجود در نمودار سمت چپ بسیار طولانی‌تر از همتایان خود در نمودار سمت راست هستند. واضح است که خط در نمودار سمت راست یک مدل پیش‌بینی بسیار بهتر از خط در نمودار سمت چپ است.

ممکن است تعجب کنید که آیا می‌توانید یک تابع ریاضی - یک تابع ضرر - ایجاد کنید که تلفات فردی را به شکل معناداری جمع کند.

Squared loss: a popular loss function

مدل‌های رگرسیون خطی که در اینجا بررسی می‌کنیم، از یک تابع ضرر به نام Squared loss (همچنین به عنوان L2 loss شناخته می‌شود) استفاده می‌کنند. مجذور ضرر برای یک مثال به صورت زیر است:

```
= the square of the difference between the label and the prediction
= (observation - prediction(x))2
= (y - y')2
```

میانگین مربعات خطا (MSE - Mean square error) میانگین تلفات مجذور در هر نمونه در کل مجموعه داده است. برای محاسبه MSE، تمام زبان‌های مجذور را برای مثال‌های جداگانه جمع کنید و سپس بر تعداد مثال‌ها تقسیم کنید:

$$MSE = \frac{1}{N} \sum_{(x,y) \in D} (y - predictions(x))^2$$

که در آن

- (x, y) نمونه‌ای است که در آن
 - x مجموعه‌ای از features (به عنوان مثال، صدای جیر جیر/دقیقه، سن، جنسیت) است که مدل از آنها برای پیش‌بینی استفاده می‌کند.
 - y label مثال است (مثلاً دما).
- $predictions(x)$ تابعی از وزن‌ها و bias‌ها در ترکیب با مجموعه features‌ها است x .
- D یک مجموعه داده حاوی نمونه‌های برچسب‌گذاری شده زیادی است که جفت (x, y) هستند.
- N تعداد نمونه‌ها در D است.

اگرچه MSE معمولاً در یادگیری ماشین استفاده می‌شود، اما نه تنها تابع ضرر عملی است و نه بهترین تابع ضرر برای همه شرایط.