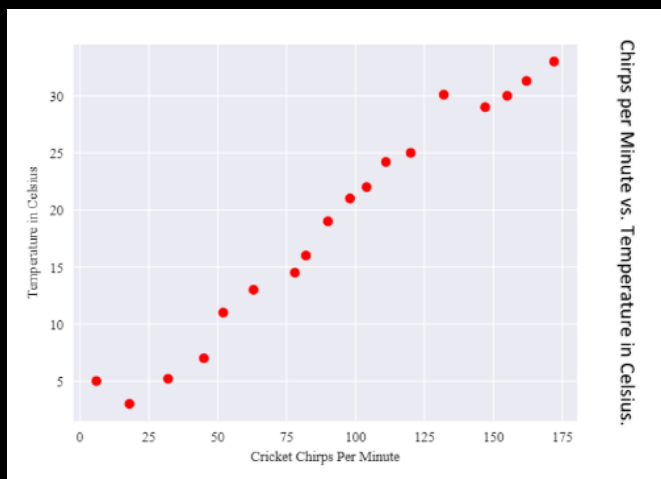


DESCENDING INTO ML

(رگرسیون خطی)

مدتهاست که مشخص شده است که جیرجیرک ها (یک گونه حشره) در روزهای گرمتر بیشتر از روزهای سردتر می کنند. برای دهه ها، دانشمندان حرفه ای و آماتور داده های مربوط به صدای جیر جیر در دقیقه و دما را فهرست بندی کرده اند. به عنوان هدیه تولد، خاله روث شما پایگاه داده کریکت خود را به شما می دهد و از شما می خواهد که مدلی برای پیش بینی این رابطه یاد بگیرید. با استفاده از این داده ها، می خواهید این رابطه را بررسی کنید.

ابتدا داده های خود را با رسم آن بررسی کنید:

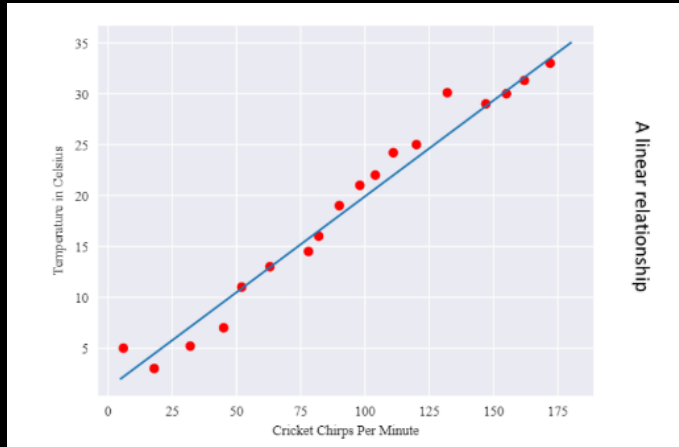


صدای جیر جیر در دقیقه در مقابل دما بر حسب سلسیوس.

همانطور که انتظار می رفت، نمودار افزایش دما را با تعداد صدای جیر جیر نشان می دهد.

آیا این رابطه بین صدا و دما خطی است؟

- بله، برای تقریب این رابطه می توانید یک خط مستقیم مانند زیر بکشید:



یک رابطه خطی

درست است، خط از هر نقطه عبور نمی کند، اما این خط به وضوح رابطه بین صدای جیر جیر و دما را نشان می دهد. با استفاده از معادله یک خط، می توانید این رابطه را به صورت زیر بنویسید:

$$y = mx + b$$

که در آن :

- y دما بر حسب سانتیگراد است—مقداری که ما سعی در پیش بینی آن داریم.
- m شیب خط است.
- x تعداد صدای جیر جیر در دقیقه است—مقدار ویژگی ورودی ما.
- b قطع y است.

طبق قرارداد در یادگیری ماشین، معادله یک مدل را کمی متفاوت می نویسید:

$$y' = b + w_1 x_1$$

که در آن:

- y' label پیش بینی شده (یک خروجی مورد نظر) است.
- b bias هست (فاصله y) است که گاهی اوقات به عنوان w_0 شناخته می شود
- w_1 وزن feature 1 است. وزن همان مفهوم "شیب" m در معادله قبلی می باشد.
- x_1 یک feature (یک ورودی شناخته شده) است.

برای استنباط (پیش‌بینی) دمای y' برای یک مقدار جدید جیک در دقیقه x ، فقط مقدار x را در این مدل جایگزین میکنیم.

اگرچه این مدل تنها از یک feature استفاده می کند، یک مدل پیچیده تر ممکن است به چندین feature متکی باشد که هر کدام وزن جداگانه ای دارند (w_1, w_2 و غیره). برای مثال، مدلی که بر سه feature متکی است ممکن است به صورت زیر باشد:

$$y' = b + w_1x_1 + w_2x_2 + w_3x_3$$