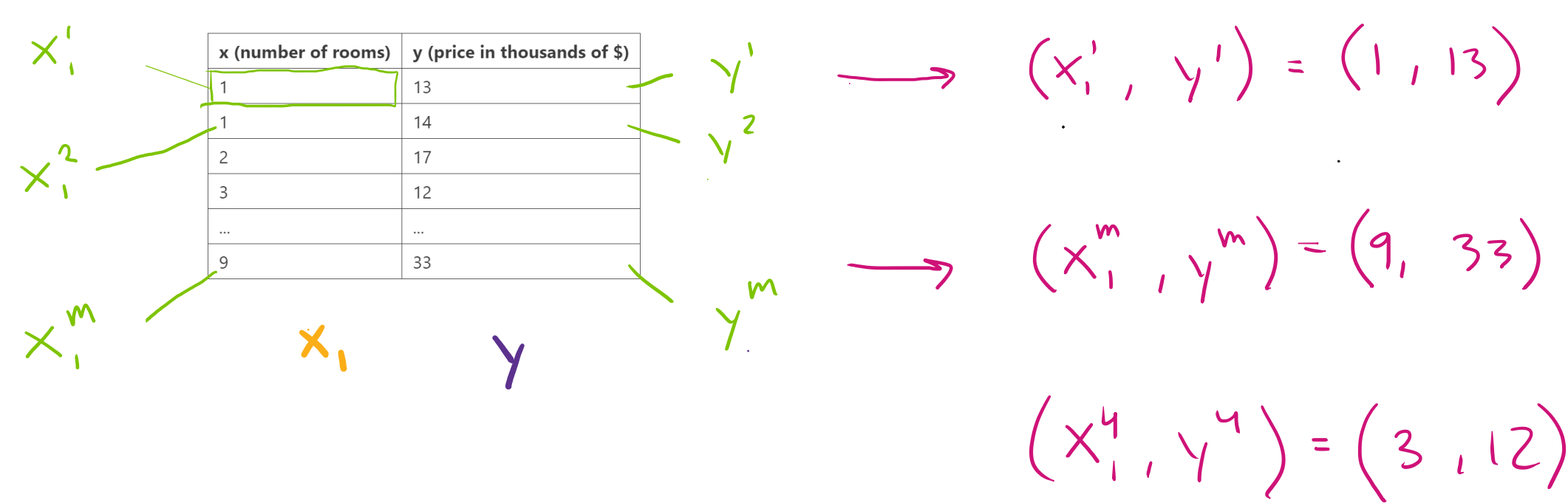


I supervised learning har vi träningsdata som består av historisk data över features respective targets

I detta fall har vi data över slutpris för bostäder (targets) givet antal rum (features) i respektive bostad



Det övergripande målet är här att "modellera" (eller att hitta ett samband mellan) $x_i \rightarrow y$

Vi kan försöka göra detta med en linjär modell

$$f_{w,b}(x_i) = w \cdot x_i + b$$

w och b är modellens parametrar

Som vi måste estimera, givet vår träningsdata

Anin att vi nu hittar värden på w och b , då kan vi använda vår modell $f_{w,b}$ för att kunna "förutspå" y för olika värden på x . Säg att vi vill predica y för x_1^2



$$f_{w,b}(x_1^2) = w \cdot x_1^2 + b = \hat{y}^2$$

Vår modells output benämner vi med \hat{y} . \hat{y} är modellens estimat av y , dvs det sannolikt

\hat{y}

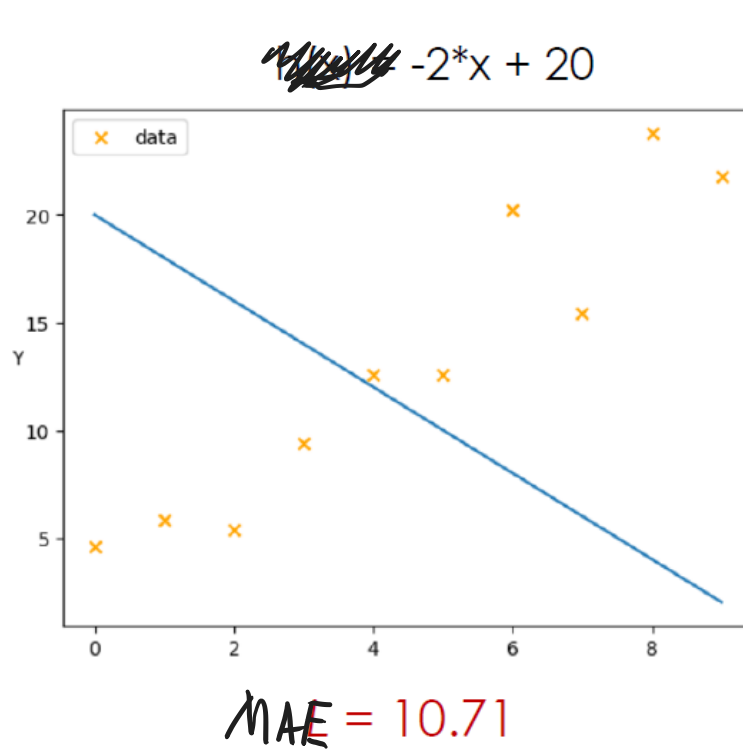
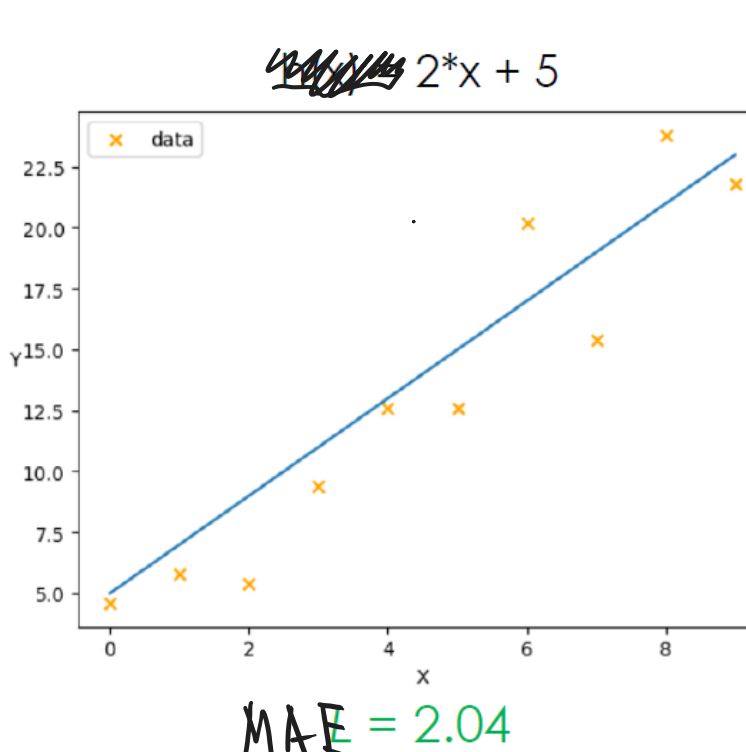
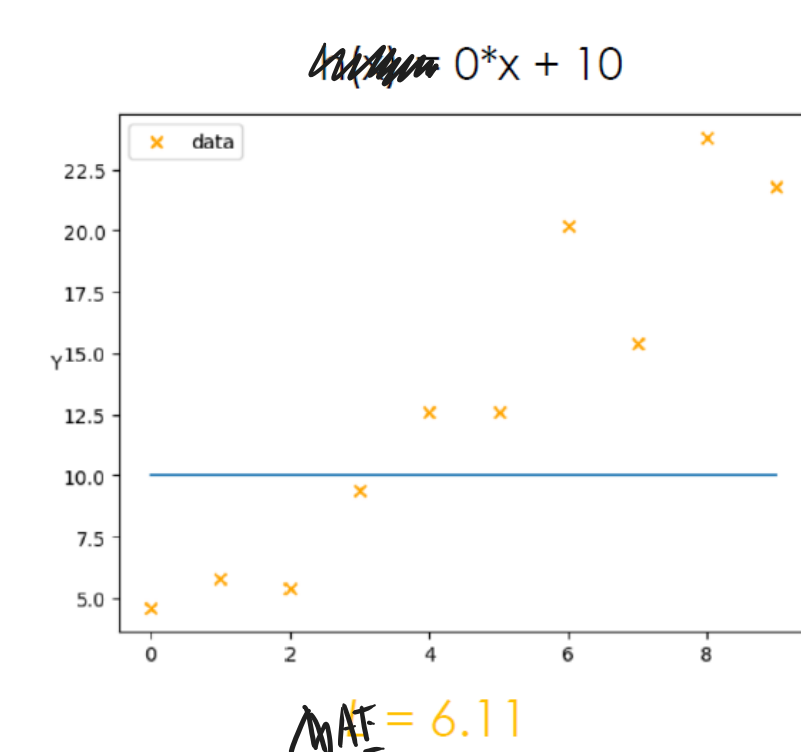
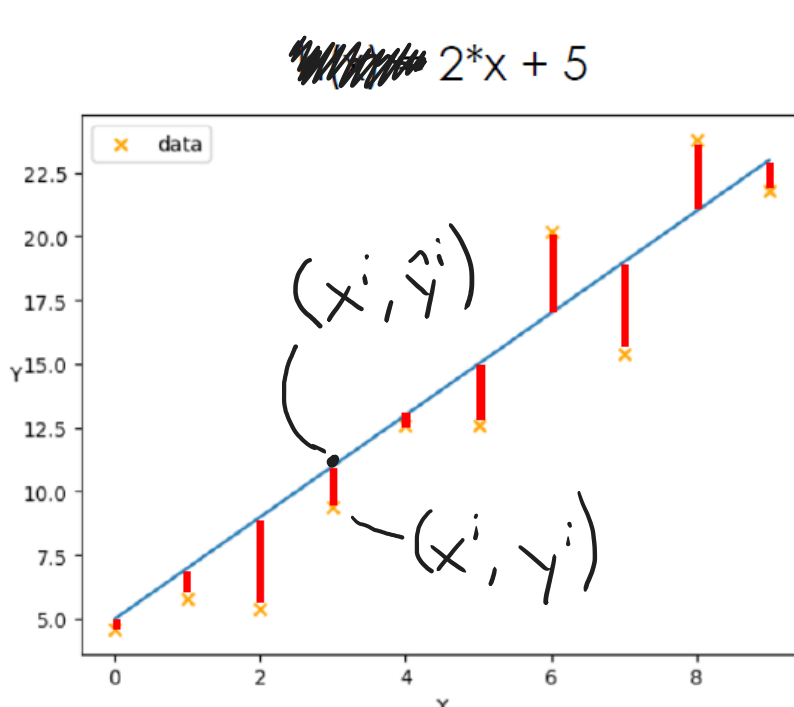
Modell Evalvering

Avståndet mellan y_i och \hat{y}_i är modellens fel i punkten x_i

$$\text{Avståndet} = |y_i - \hat{y}_i|$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

↑
medelabsolut felet / mean absolute error



1. Fit your model, dvs hitta w och b
2. Beräkna \hat{y}_i för alla x_i i din träningsdata
3. Jämför med ground truth, dvs beräkna MAE
4. Lågt MAE \leftrightarrow Bättre modell *

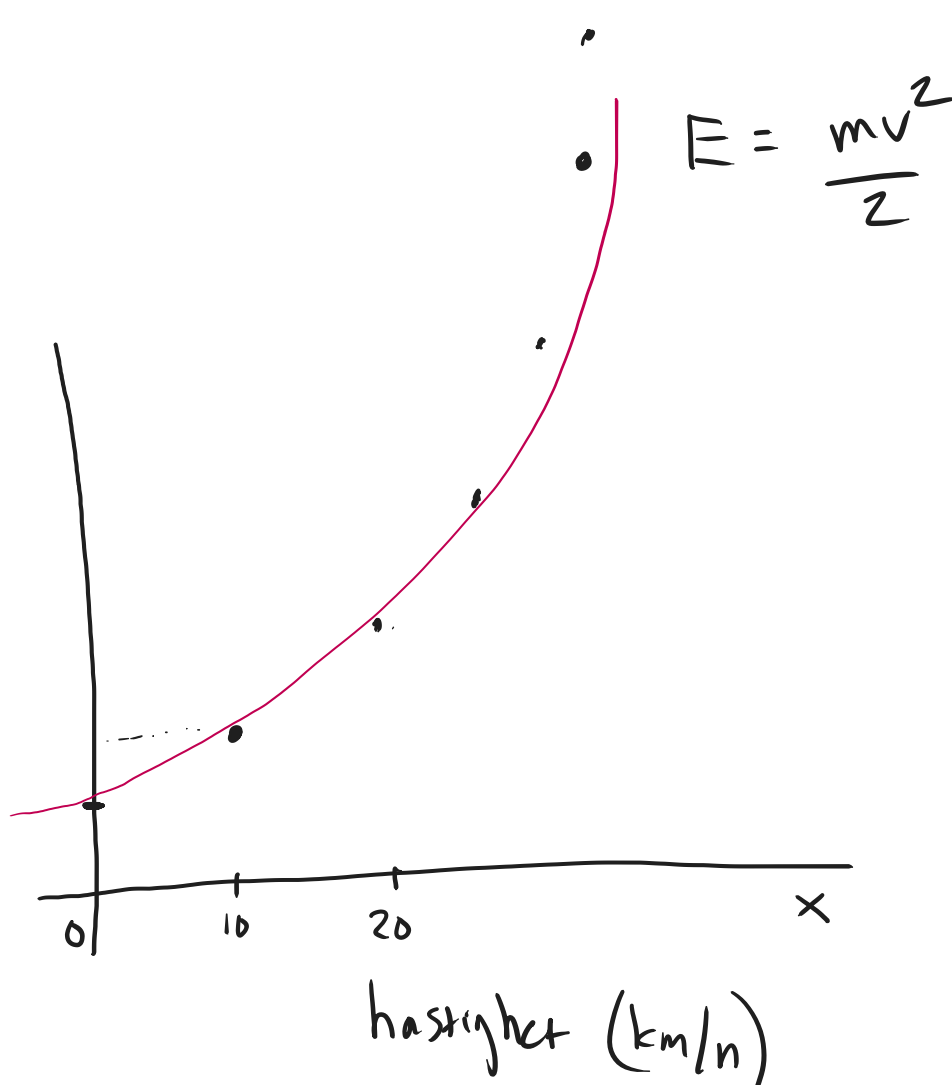
Polynomial Regression

polynom: $y = ax^2 + bx + c$

$$y = ax^2 + bx + c$$

$$y = ax + b$$

(kcal/h)



kcal	km/h	(km/h) ²
100	0	0
300	10	100
...
10000	100	10000

↑ ↑

target y feature x

Vi ser att en linjär modell

$$f_{w,b}(x) = w \cdot x + b \text{ inte passar bra här}$$

Istället kan vi testa en polynom modell av graden 2

$$f_{w,b}(x) = w_1 x_1 + w_2 (x_1)^2 + b$$

Uppgift: hitta värden på w_1, w_2 och b som minimerar MAE