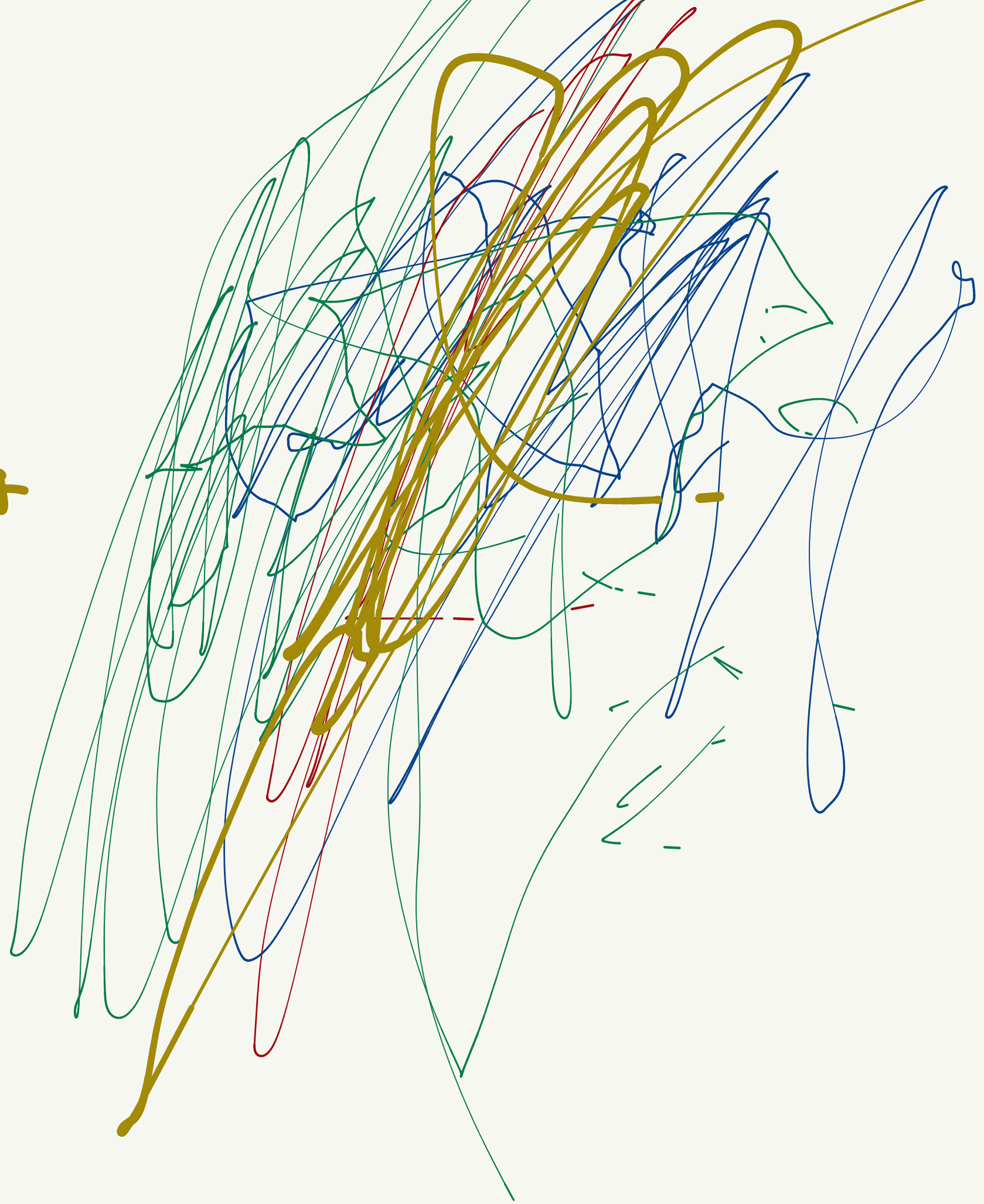


$$y = f(x) +$$



# 1. Sigmoidfunksjonen

Eks. Temperatur.

## 2. Utvide til $z = \beta_0 + \beta_1 x_1$

- Regne på hvordan  $z: f(z) = \frac{1}{2}$  endrer seg med  $\beta_0$  og  $\beta_1$

Justerer av sigmoidfunksjonen

## 3. Funktionsfitting av sigmoidfunksjonen til datasettet fra sist uke

Logistisk regresjon.

## 4. Flere variable, kontinuerlige og binære $x_i$ .

Balanse + student? Konseptuelt med underoppgaver.

## 5. Funktionsfitting m/ flere variable. Kredittkort: (balanse, student?)

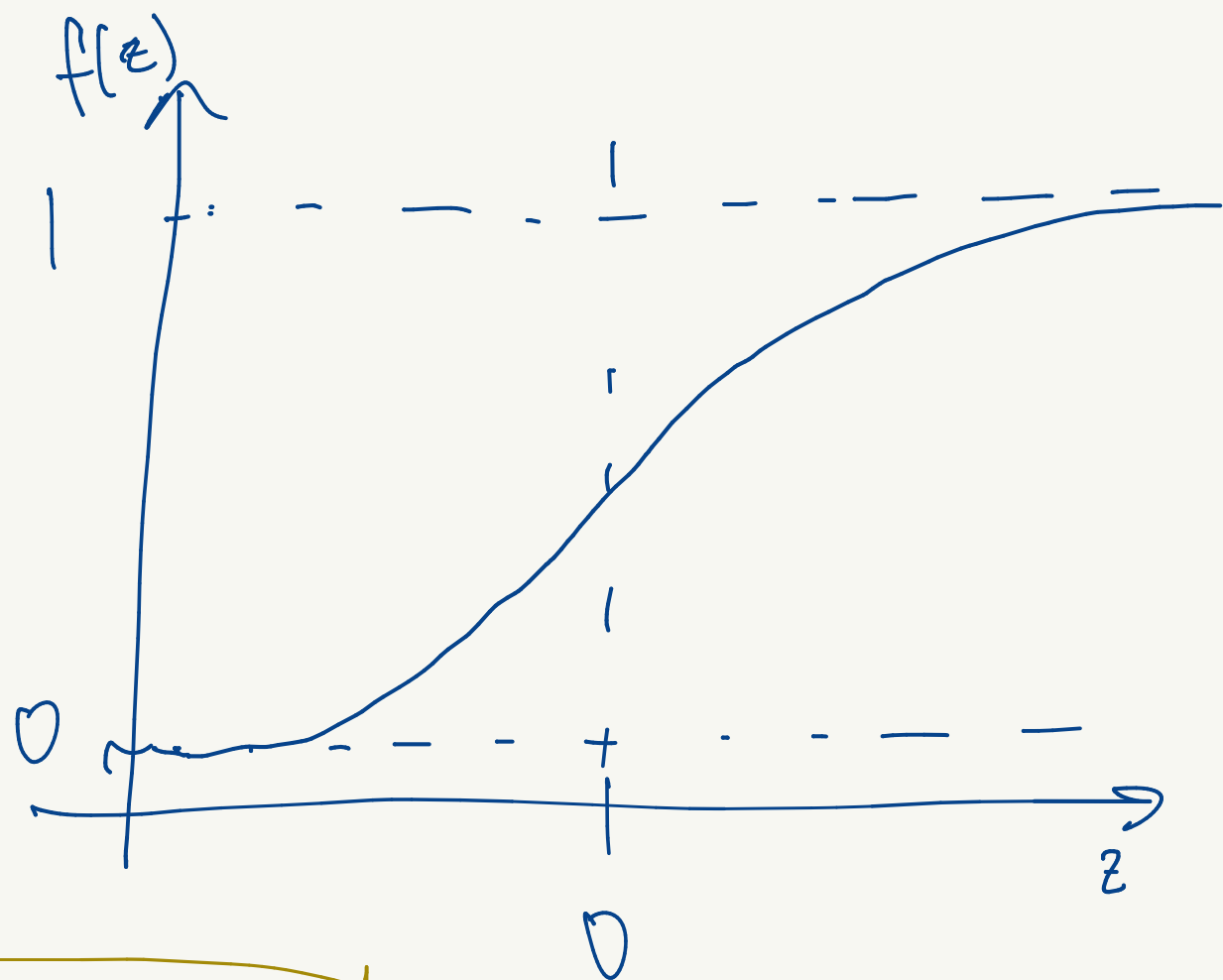
1 jupyter.

## 6. Begynne på Titanic-obliggen.

Vis hele oppg. 1.

1)

# Sigmoidfunktioner.



$$f(z) = \frac{e^z}{1 + e^z}$$

Matematisk søv -

$$f(-\infty) = 0 \quad \text{fra } 0$$

$$f(\infty) = 1 \quad \text{til } 1$$

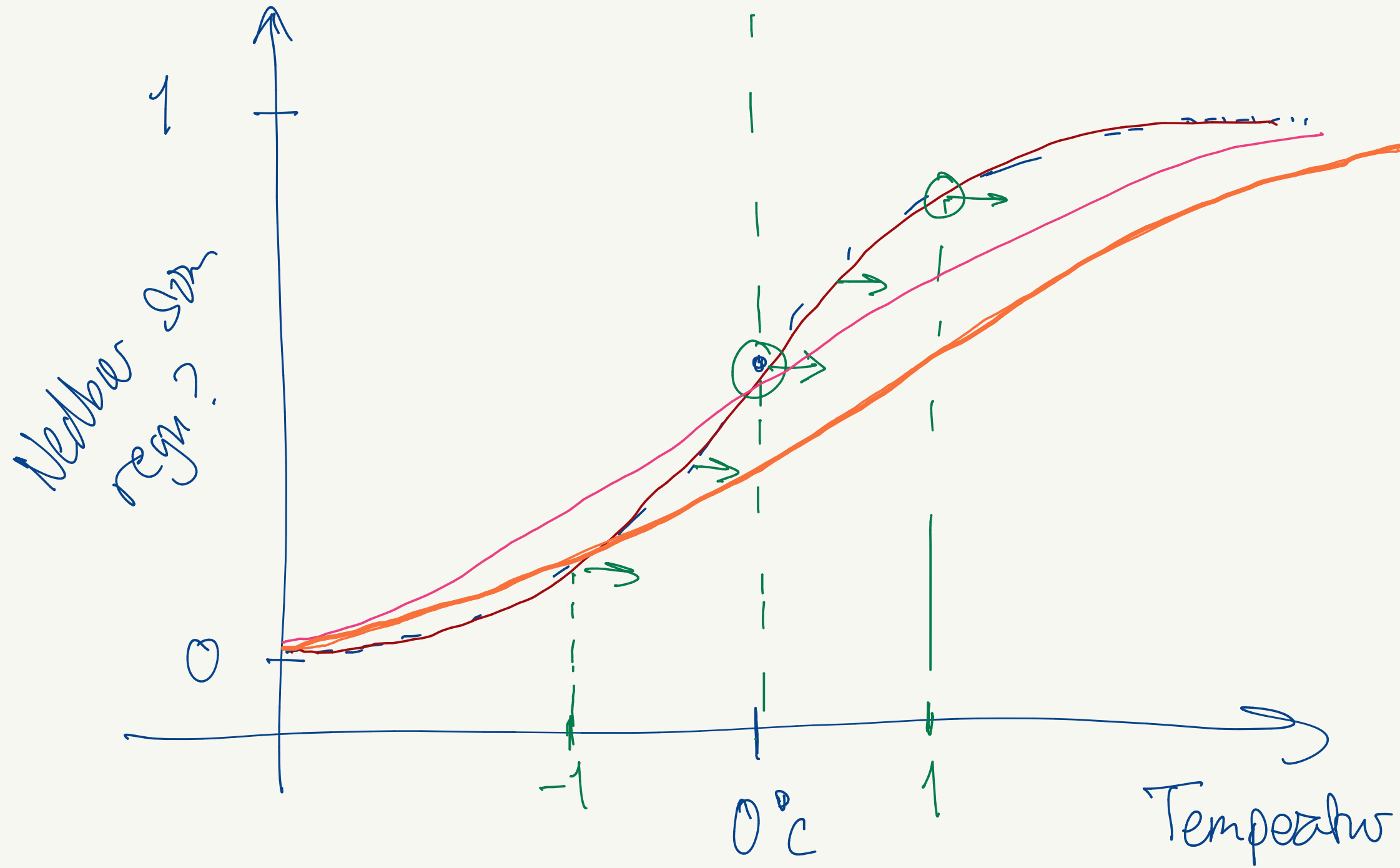
Modell for f. eks. "Kommer nedbøren til å falle som regn?"

1 (Ja) for store T og 0 (Nei) for små T. og 50/50 ved 0.

MEN: Hva om vi vil flytte  $f = \frac{1}{2}$  ved 0?

# Kort eksempel

- 1) flytter temperaturen for 50/50
- 2) Reduser sensitivitet





2) Vi ser

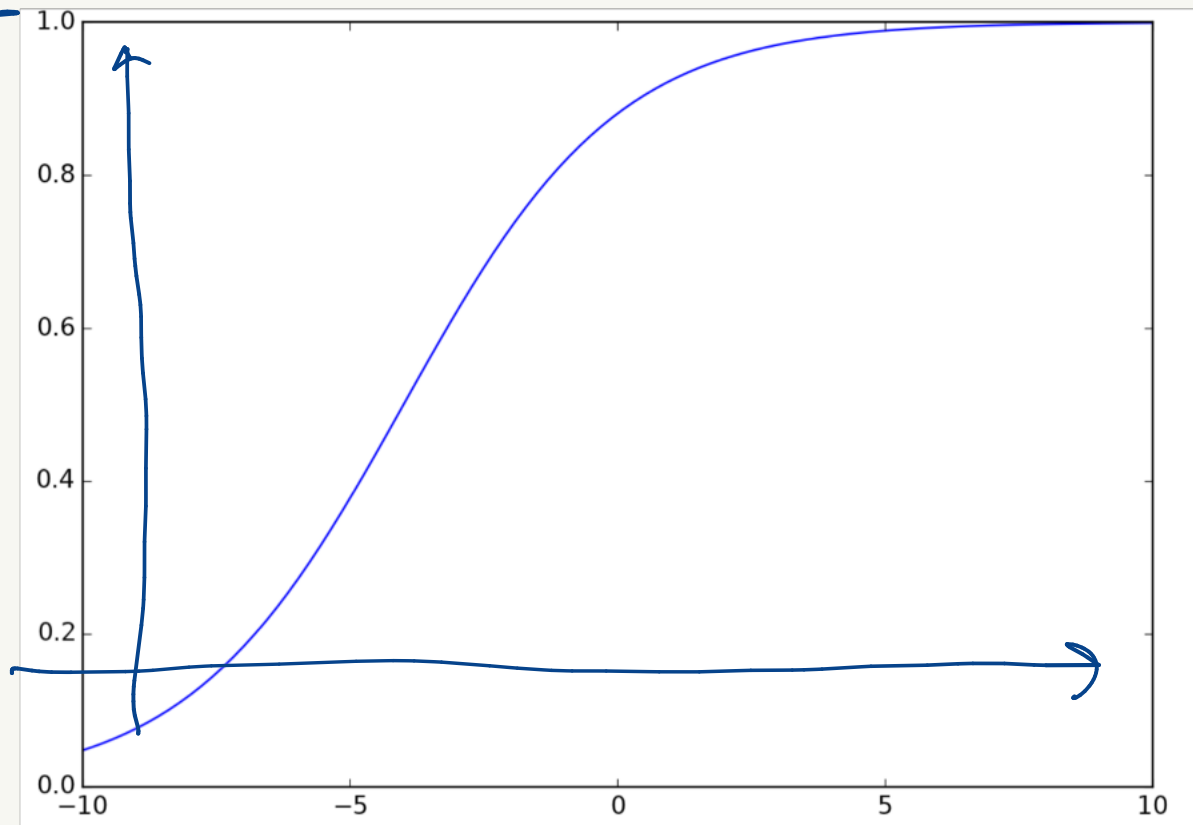
$$Z = \beta_0 + \beta_1 x_1$$



Parametre i modellen var

La oss prøve  $\beta_0 = 2$  og  $\beta_1 = 0,5$

Plot



Litt algebra for å kvantifisere.

$$f(x_1) = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}}$$

Om vi ønsker å bestemme  
hvor vendepunktet skal være:

Løse for

$$f(x) = \frac{1}{2}, \text{ hva er da } x?$$

(Neste side).

$$\frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}} = \frac{1}{2} \rightarrow e^{\beta_0 + \beta_1 x_1} = 1$$

$$\frac{z}{1+z} = \frac{1}{2}$$

$$z = \frac{1}{2} + \frac{z}{2}$$

$$\frac{z}{2} = \frac{1}{2} \Rightarrow \underline{\underline{z = 1}}$$

$$e^{\beta_0 + \beta_1 x_1} = 1$$

$$\ln(e^{\beta_0 + \beta_1 x_1}) = \ln(1)$$

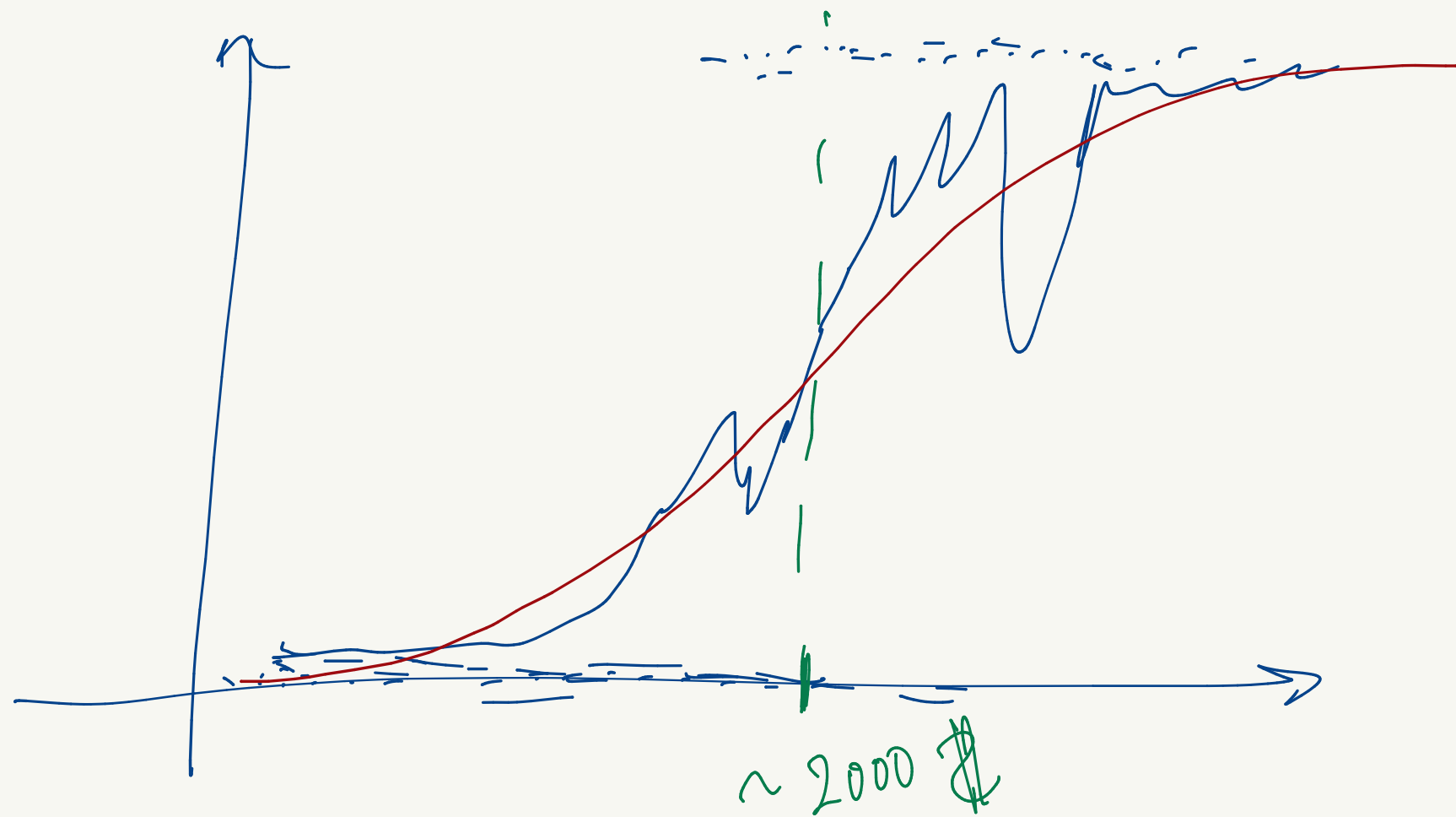
$$\beta_0 + \beta_1 x_1 = 0$$

$$\beta_1 x_1 = -\beta_0$$

$$x_1 = \frac{-\beta_0}{\beta_1}$$

for at  $f(x_1)$  skal være  $\frac{1}{2}$   
ved  $x_1$

Förrige programmeringsåret så vi på:



- 1) Koordinatsystem
- 2) Data-punkter
- 3) Snitt av punkter, lignesigmoid
- 4) Sigmoid.
- 5) Markve ~ 2000 \$

La oss programmera detta med pseudo-kode!

# Manuell regression

1 def logistiz (X,  $\beta_0$ ,  $\beta_1$ ):

2 return  $\frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$

3 X\_vende = 2000

4  $\beta_1 = 0,01$  # (eller noe)

5  $\beta_0 = -\beta_1 X_{vende}$ .

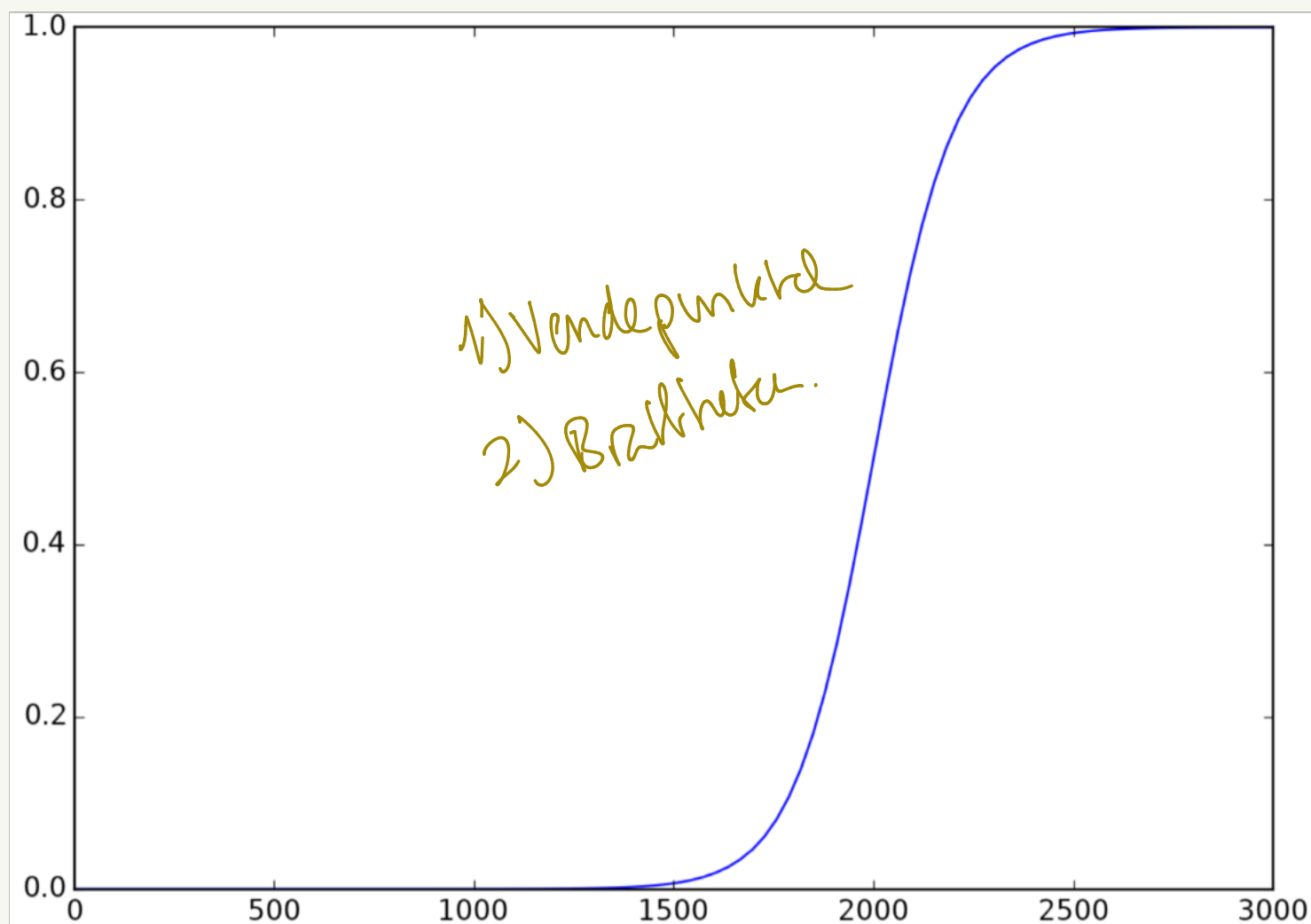
6 X = linspace (0, 3000, 100)

7 y = logistiz (X,  $\beta_0$ ,  $\beta_1$ )

8 plot (X, y)

Mellomregning

$$X_1 = -\frac{\beta_0}{\beta_1} \Rightarrow \beta_0 = -\beta_1 X_1$$



Dette var manuell trening  
av modellen, la oss gå over til  
automatisert løsning.

Live - programmer

logistiske regresjon for  
kredittkonfeksenspellet.

<1>

Bedre å følge med enn å kode med sikkerhet her.

Notes  
→

# Eller vise løsninger

```
# antar vi her beregnet andel mislighold fra før
from sklearn.linear_model import LogisticRegression

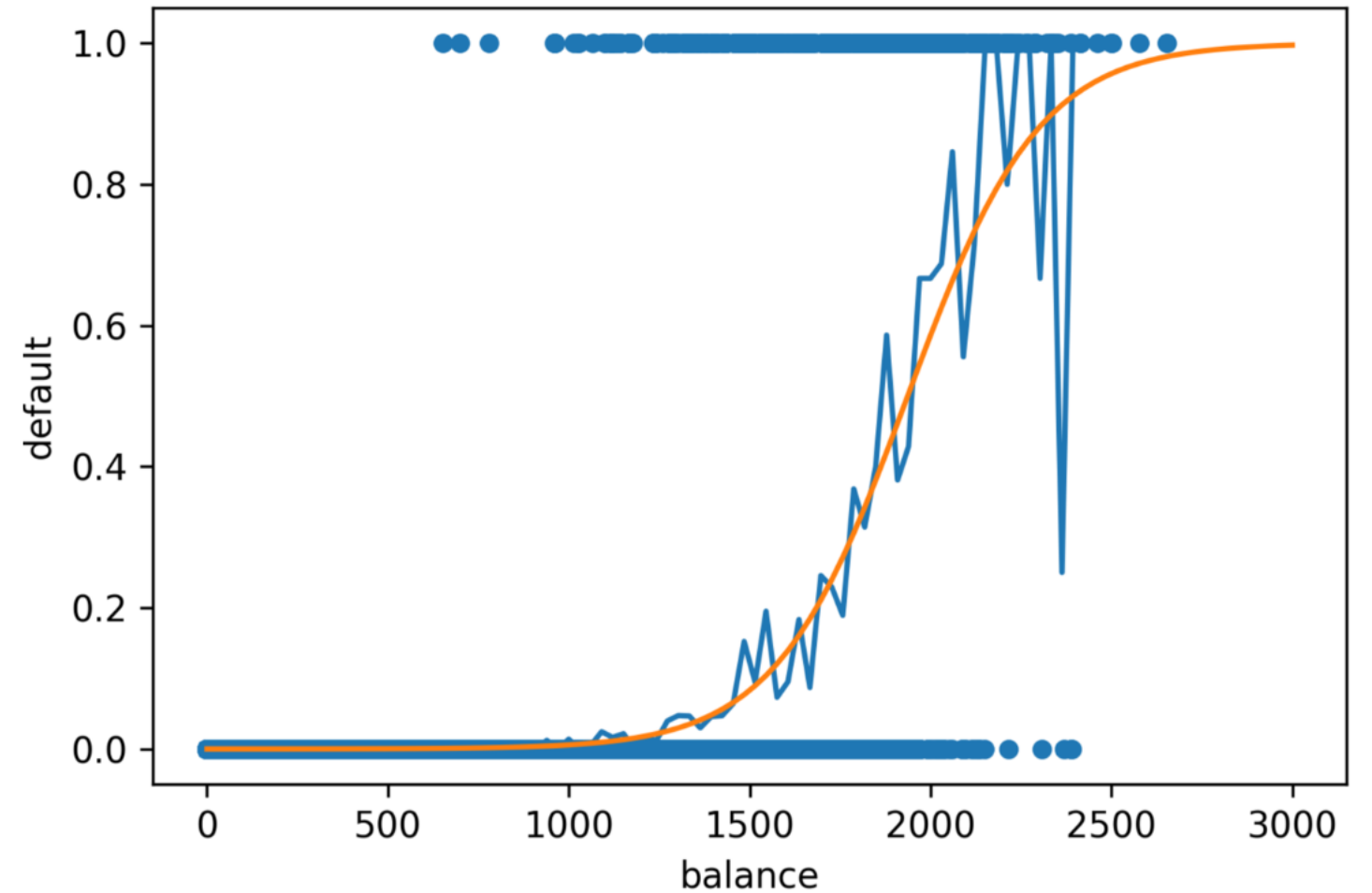
model = LogisticRegression()
fit = model.fit(data[["balance"]], data["default"])

print("Coefficients: ", model.coef_)
print("Intercept: ", model.intercept_)

beta0 = model.intercept_[0]
beta1 = model.coef_[0][0]

def sigmoid(x, beta0, beta1):
    z = beta0 + beta1 * x
    return np.exp(z) / (1 + np.exp(z))

data.plot.scatter("balance", "default")
plt.plot(x[:-1], andel_mislighold)
plt.plot(x, sigmoid(x, beta0, beta1))
plt.savefig("default.png", dpi=300)
```



# To input-variable

F. eks både kreditbalance OG studentstatus.

Forkalt:

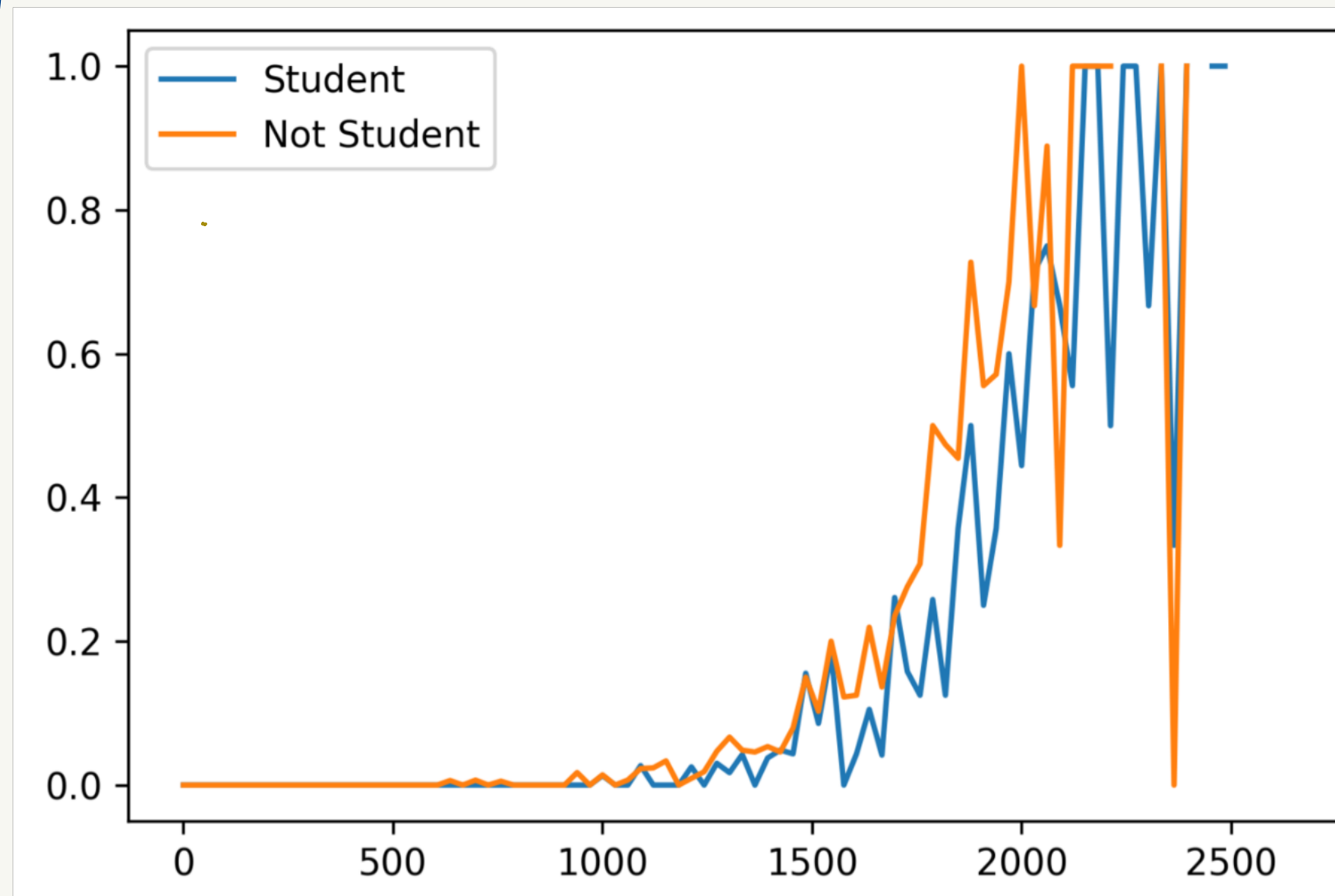
$$f(z) = \frac{e^z}{1 + e^z}$$

MEN:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Vendepunkt  $\beta_0$   
Balans Kontinuerlig variabel  $x_1$   
Student? Binær variabel  $x_2$

Datasættet vort





# Underveis oppgave

• Er studenter, basert på dataene presentert her, bedre eller

dårligere betalere

enn andre folk?

Tenke i 30 sekunder selv.

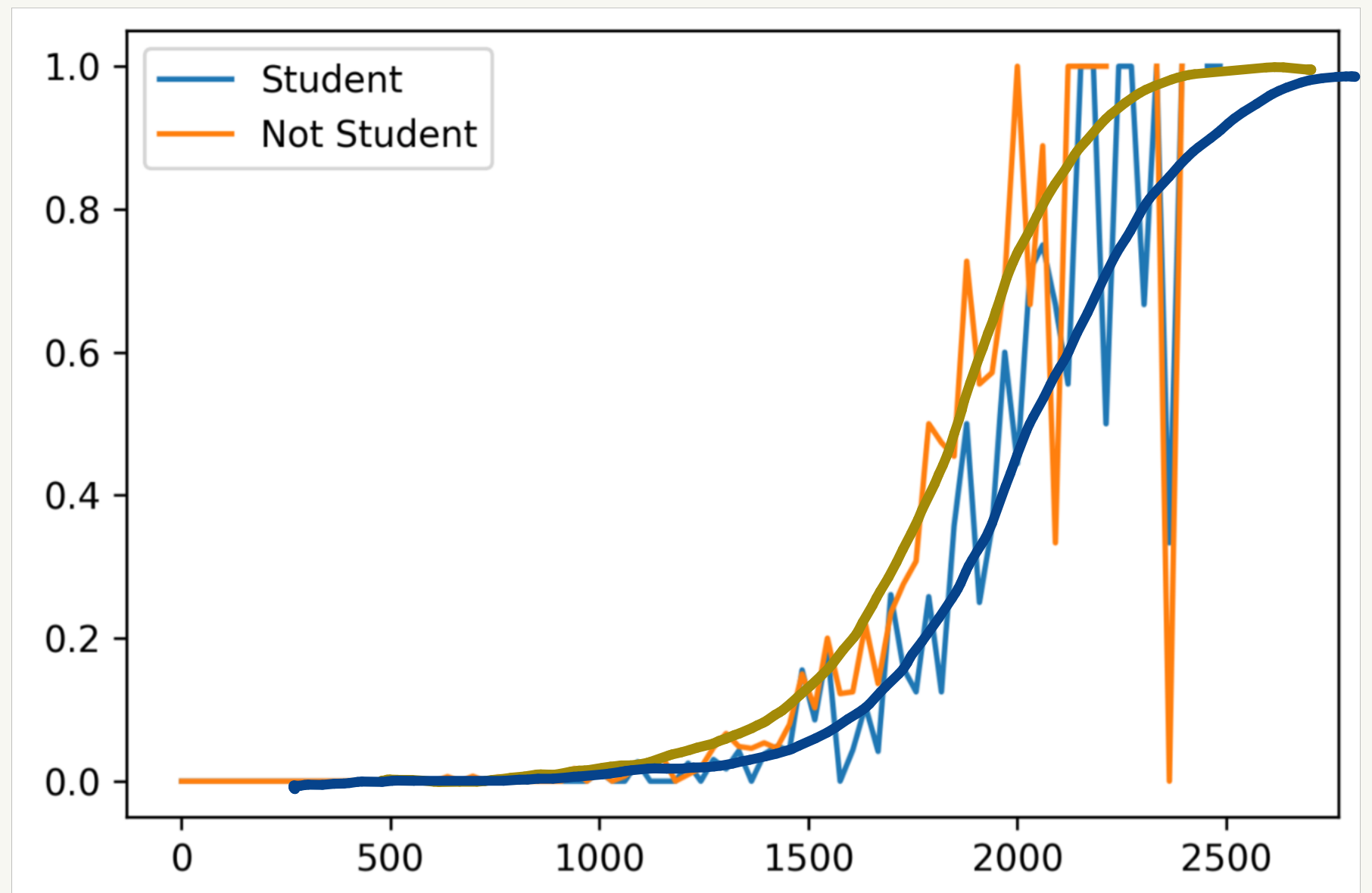


Bedre



Dårligere

1) Vi ønsker å lage en modell der  $\mathbb{R}^2 \times \mathbb{R}^2$  forskyver grafen.



Live - programmer

Modell med fløe input -

variable: balance og studentstatus

Følge med, ikke kode med.

{2}

# Tolkning av koefficienter

model. intercept  $\rightarrow [\beta_0]$

sensitivitet för saldo

model. coef  $\rightarrow [\beta_1, \beta_2]$

sensitivitet för studentstatus.

i likningen

$$\frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}$$

$\beta_0 = -10,74 \dots$  - Flykter på vändpunkten.

$\beta_1 = 0,0057 \dots$  - ganges med fall  $\sim 1000 \text{ \$}$

$\beta_2 = -0,699 \dots$  - ganges med fall  $\sim 1$  (Ja/Nei). Denne er  $\sim 6-7\%$  av  $\beta_0$

# Flere variabler

Det er ikke grense...

$$e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \dots}$$

$$1 + e^{\beta_0 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \dots}$$

Dette gir de: Titanz - obliger -

Et DBS om  
numerisk ustabilitet.

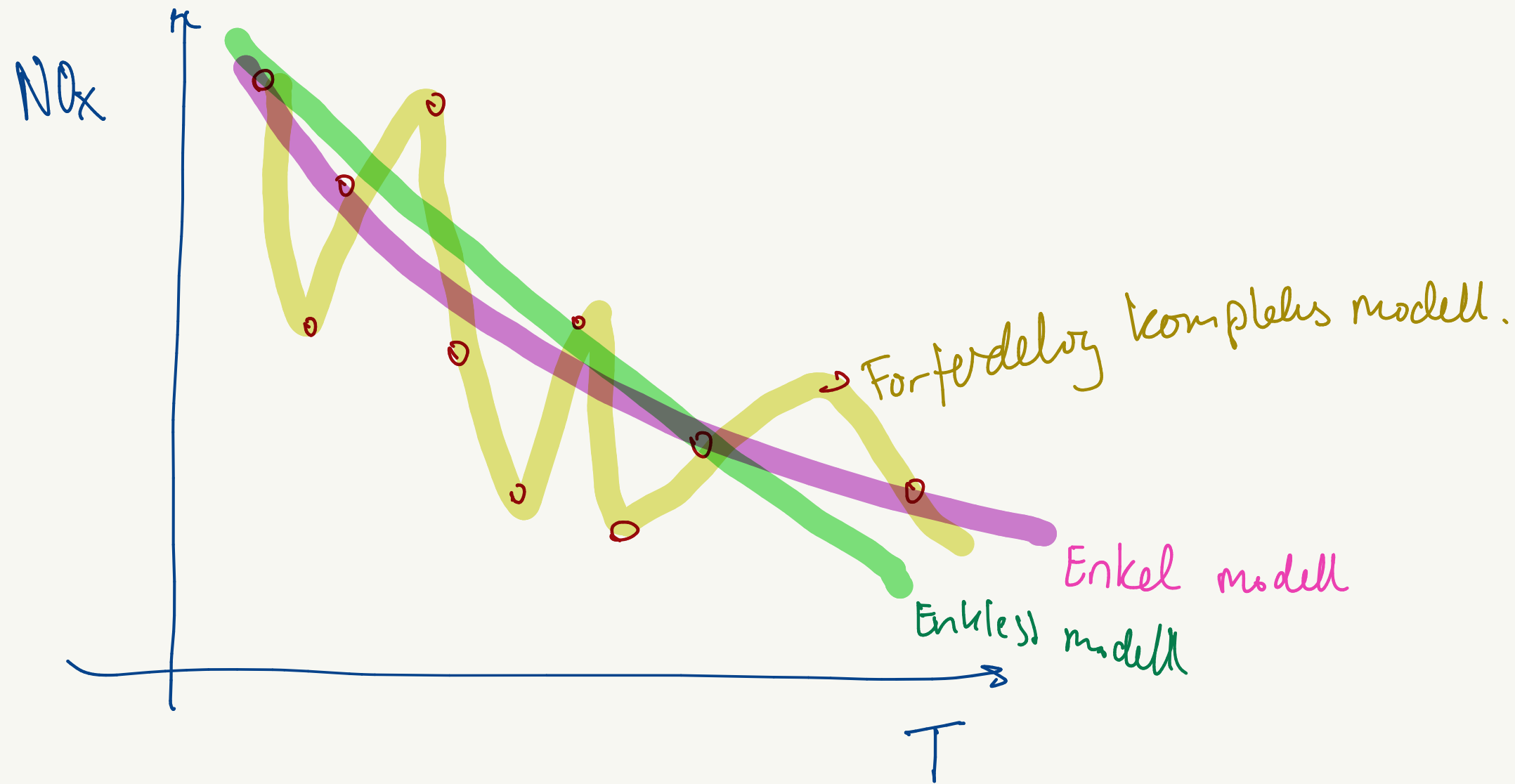
$e^{\text{Stort}}$  = veldig stort

Kan gi overflow.

Derfor  $\frac{1}{1+e^{-x}}$  for store  
tall.

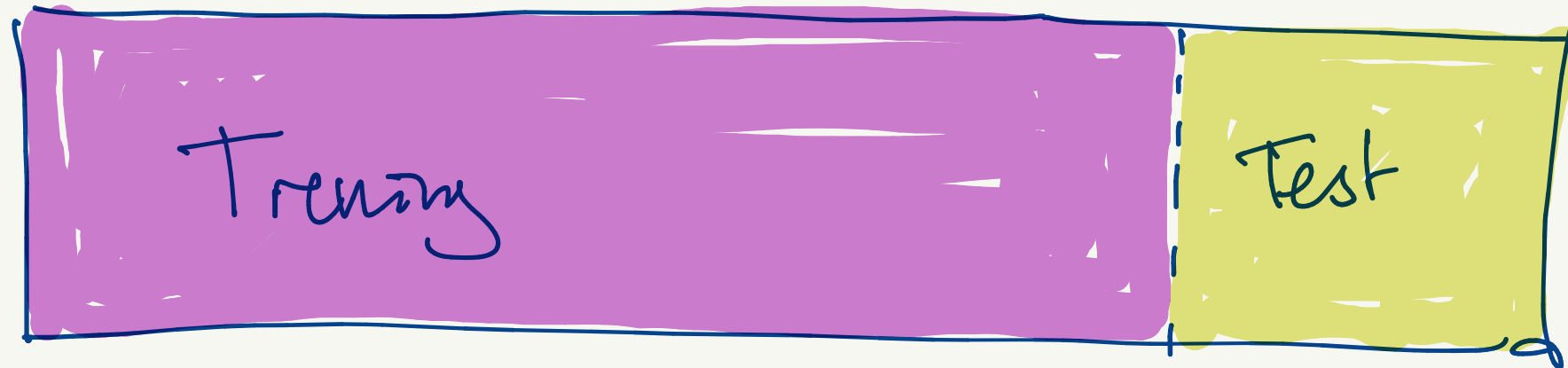
# Illustrasjon av overfitting -

(om deler blot for mange variable).

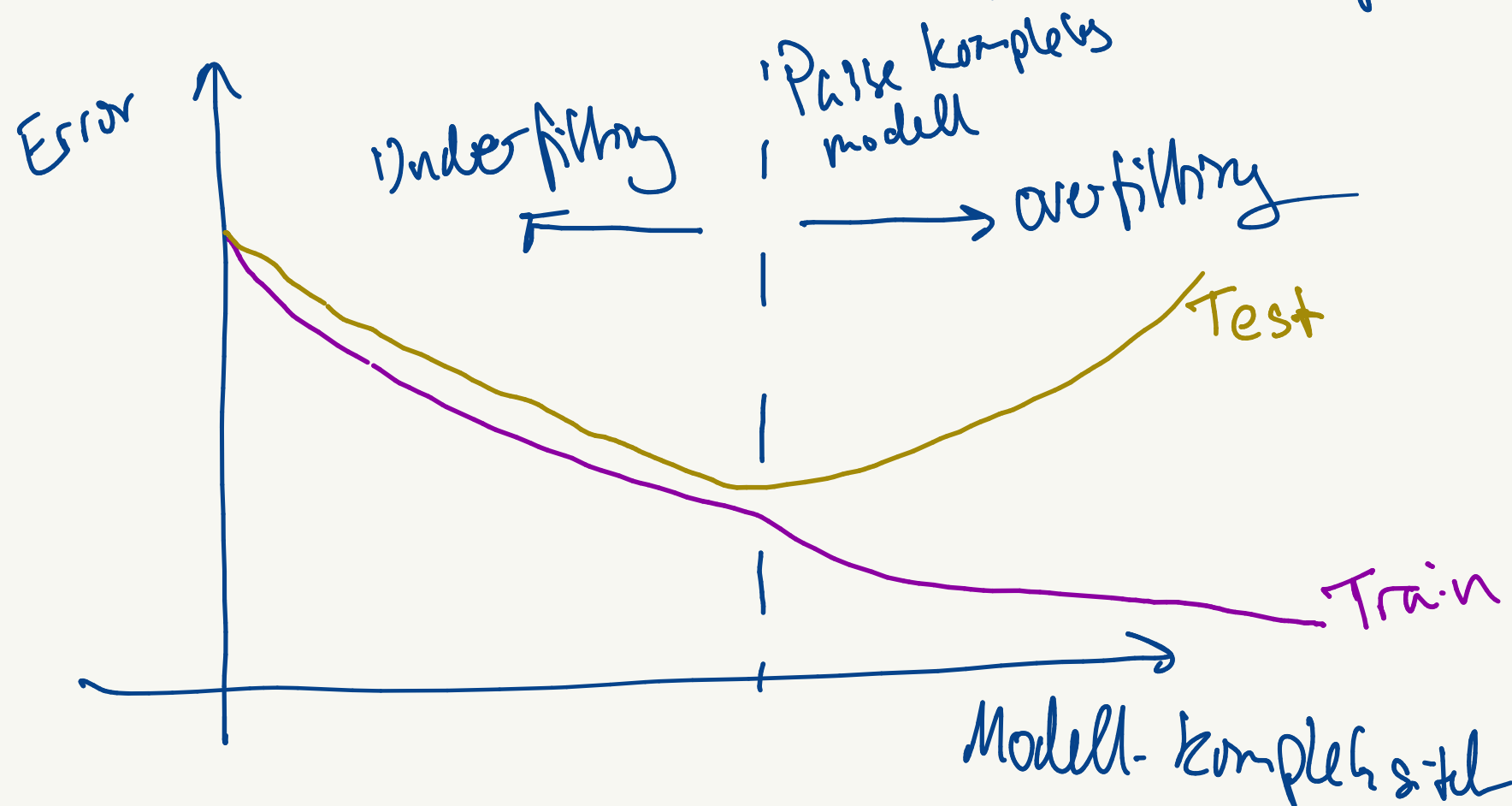


Hvordan kan vi vite at modeller er god for andre data enn de vi har trent modeller med?

Data



Train-Test Split



Vi trener så modeller KUN med treningsdata, og bruker test-data til å evaluere modellen.

# One hot encoding

Dyr	Levelde	BMI
Gris	3	82
Ku	8	71
Gris	2	94
Sam	9	60
Sam	12	43
Ku	9	90

↑ like tall.  
↑ tall  
↑ tall

~~Ku: 1~~

~~Gris: 2~~

~~Sam: 3~~

~~$\Rightarrow \frac{Ku + Sam}{2} = \text{gris}$~~

!!  
..

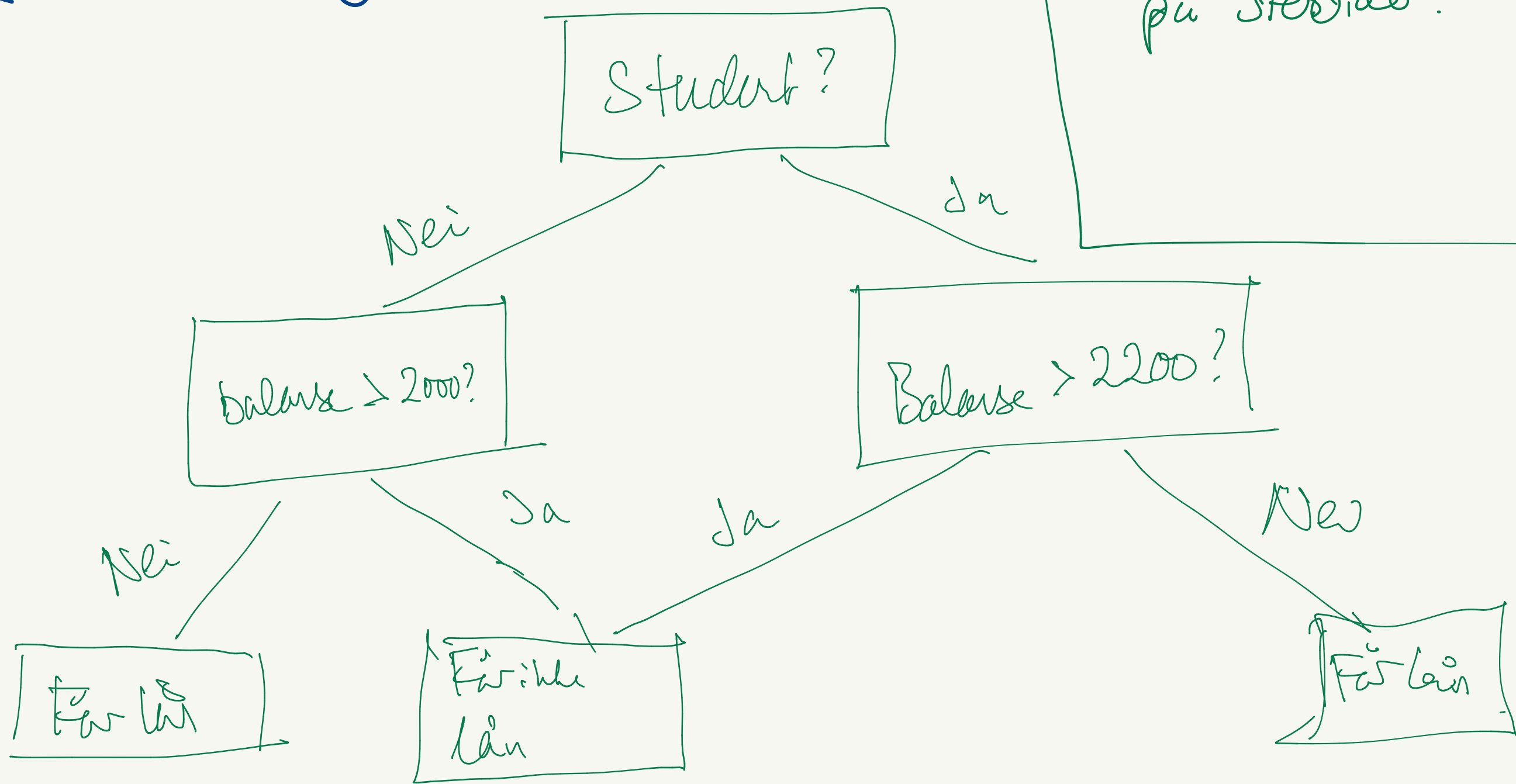
## One hot encoding

Ku?	Gris?	Sam?	Levelde	BMI
0	1	0	3	82
1	0	0	8	71
0	1	0	2	94
0	0	1	9	60
0	0	1	12	43
1	0	0	9	90

get-dummies : pandas.



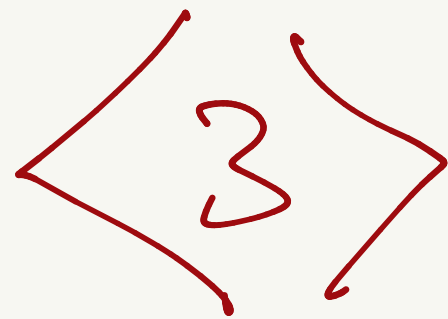
# Beslutnings træ



XGBoost er beslutnings træ på stovides.

Scikit-learn lar oss gjøre dette automatisk  
Mange træer blir en skog.

Live-kode  
et beslutningsre.



# De første stegene: Titanic - obligen

Viktighet

1) Lese inn data (1a)

2) Preparere data inkl. train, test-split (1b-d)

3) Trene logistiske regresjonsmodell. (2b)

4) Teste logistiske regresjonsmodell på test-data. (2c)

5) Sette opp beslutningstrær i scikit-learn (3a)

6) Teste beslutningstrær (3b) *Uttfordring*

XGBoost

Frivillig.

## De senere stegene

1) Lese data

2) Preparere data

3) Trene modell

4) Evaluere modell

TEMPLATE

- Kronidder inne?

- status oblig ?

- Oraltimer -

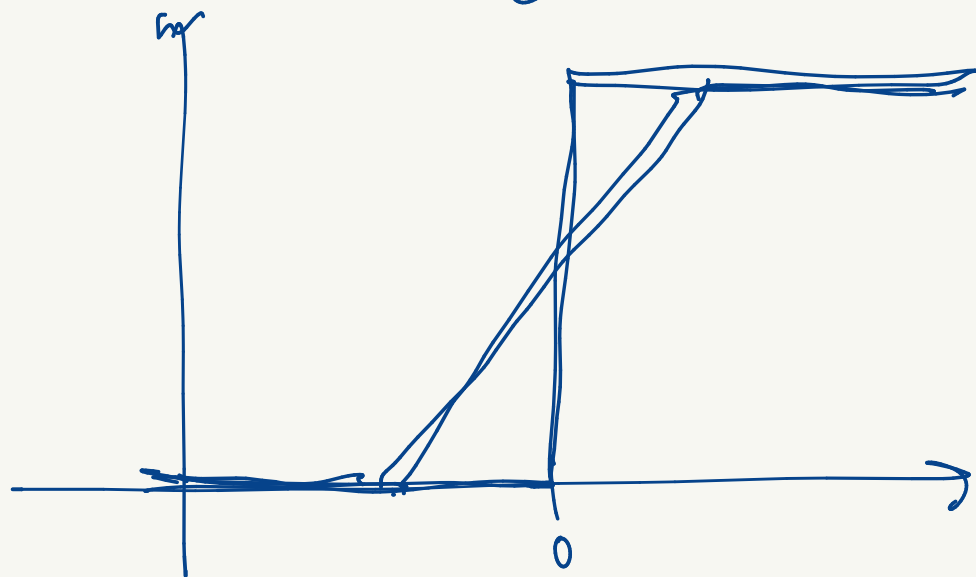
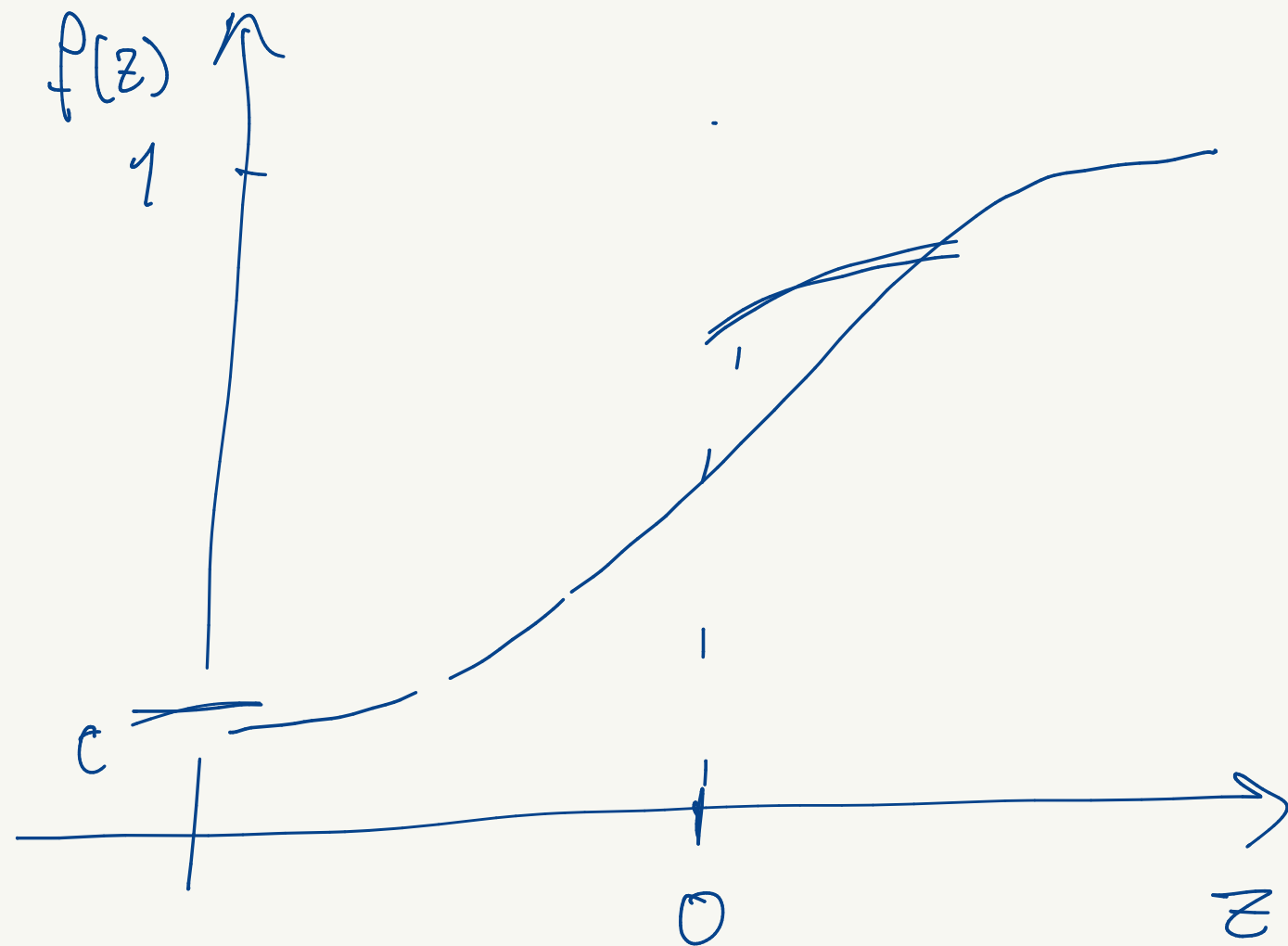
T 14-16

F 10-12

} Legges ut på carvas.

1. Sigmoid funksjonen
2. Utvide til  $z = \beta_0 + \beta_1 x_1$   
Regne på  $z$ :  $f(z) = \frac{1}{2}$
3. Funksjons tilpassing av sigmoidfunksjonen til datasettet fra sist uke
4. Flere variable, kontinuerlige og binære  $x_i$ .
5. Funksjons tilpassing m/ flere variable. Kredittkort: (balanse, student?)
6. Begynne på Titanic-obehgen.

# Sigmoid funktionen



$$f(\infty) = 1$$

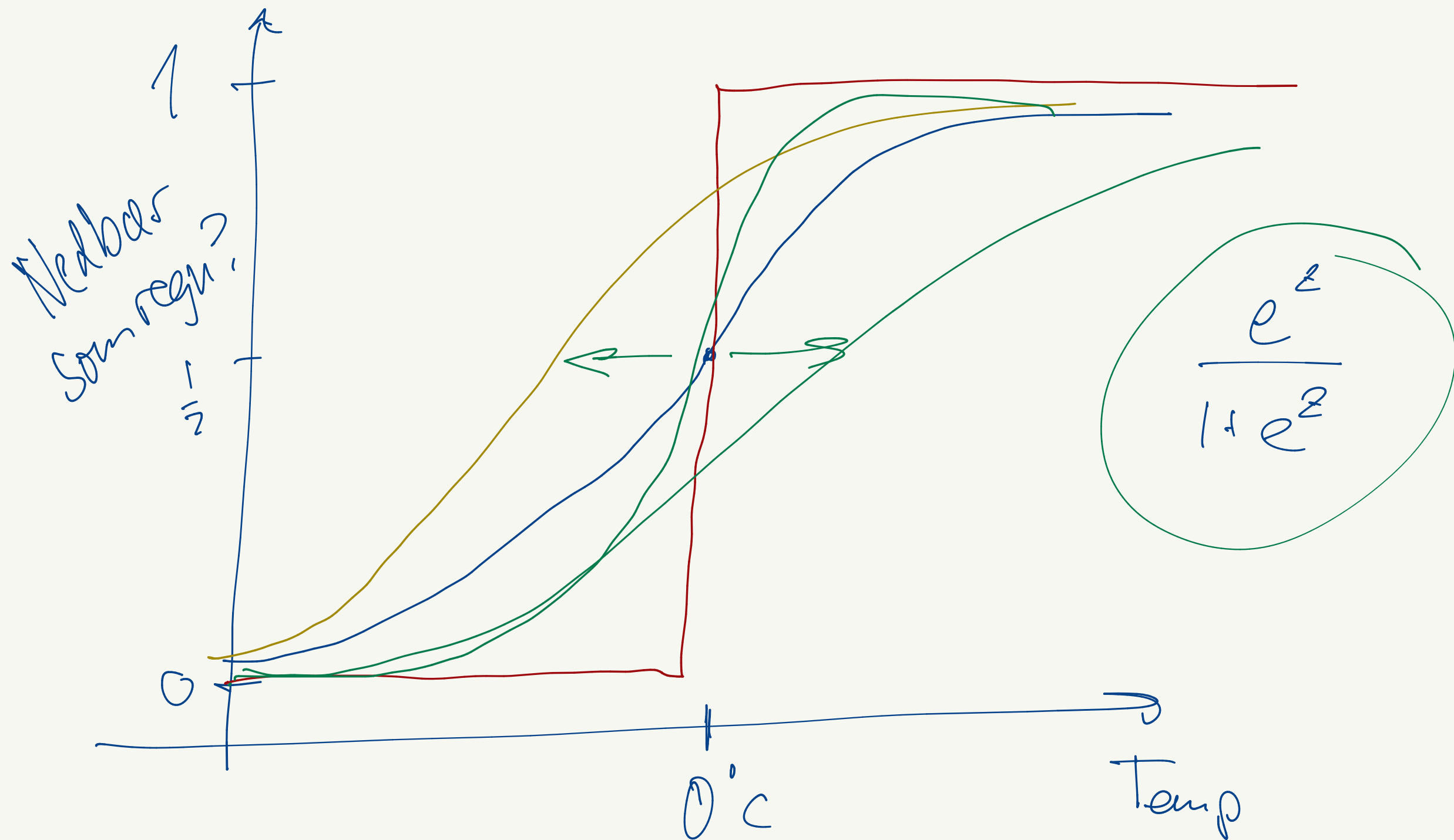
$$f(-\infty) = 0$$

$$f(z) = \frac{e^z}{1 + e^z}$$

Sigmoidfunktion



# Eksempel



$\beta_0$  og  $\beta_1$

$$f(z) = f(x_i, \beta_0, \beta_1)$$

$$f(z) = \frac{e^z}{1 + e^z}$$

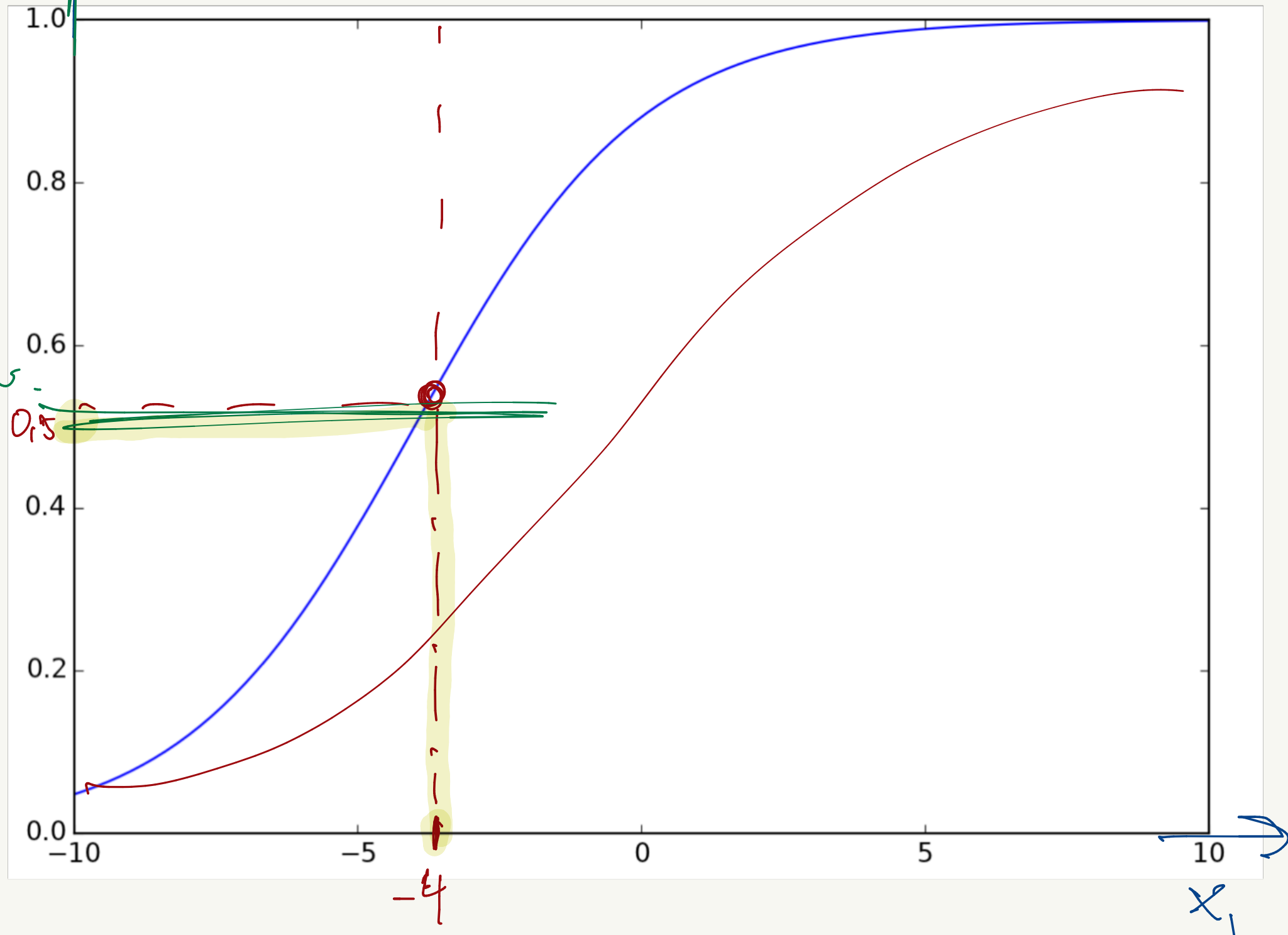
$$z = \beta_0 + \beta_1 \cdot x_i$$

Modellparametre

Temperatur

$$\beta_0 = 2$$

$$\beta_1 = 0,5$$



(Finne  $x_1$ :  $f(x_1, \beta_0, \beta_1) = \frac{1}{2}$ )  $\frac{e^z}{1+e^z}$ ,  $z = \beta_0 + \beta_1 x_1$

$$f(x_1, \beta_0, \beta_1) = \frac{e^{\beta_0 + \beta_1 x_1}}{1 + e^{\beta_0 + \beta_1 x_1}} \stackrel{!}{=} \frac{1}{2}$$

$$e^z = u$$

$$\frac{e^z}{1+e^z} = \frac{u}{1+u} \stackrel{!}{=} \frac{1}{2} \quad | \cdot (1+u)$$

$$u = \frac{1}{2} (1+u)$$

$$u = \frac{1}{2} + \frac{u}{2} \quad | - \frac{u}{2}$$

$$\frac{u}{2} = \frac{1}{2} \Rightarrow \boxed{u = 1}$$

$$u = e^z$$

$$e^z = 1$$

$$e^{\beta_0 + \beta_1 x_1} = 1$$

$$\ln(e^{\beta_0 + \beta_1 x_1}) = \ln(1)$$

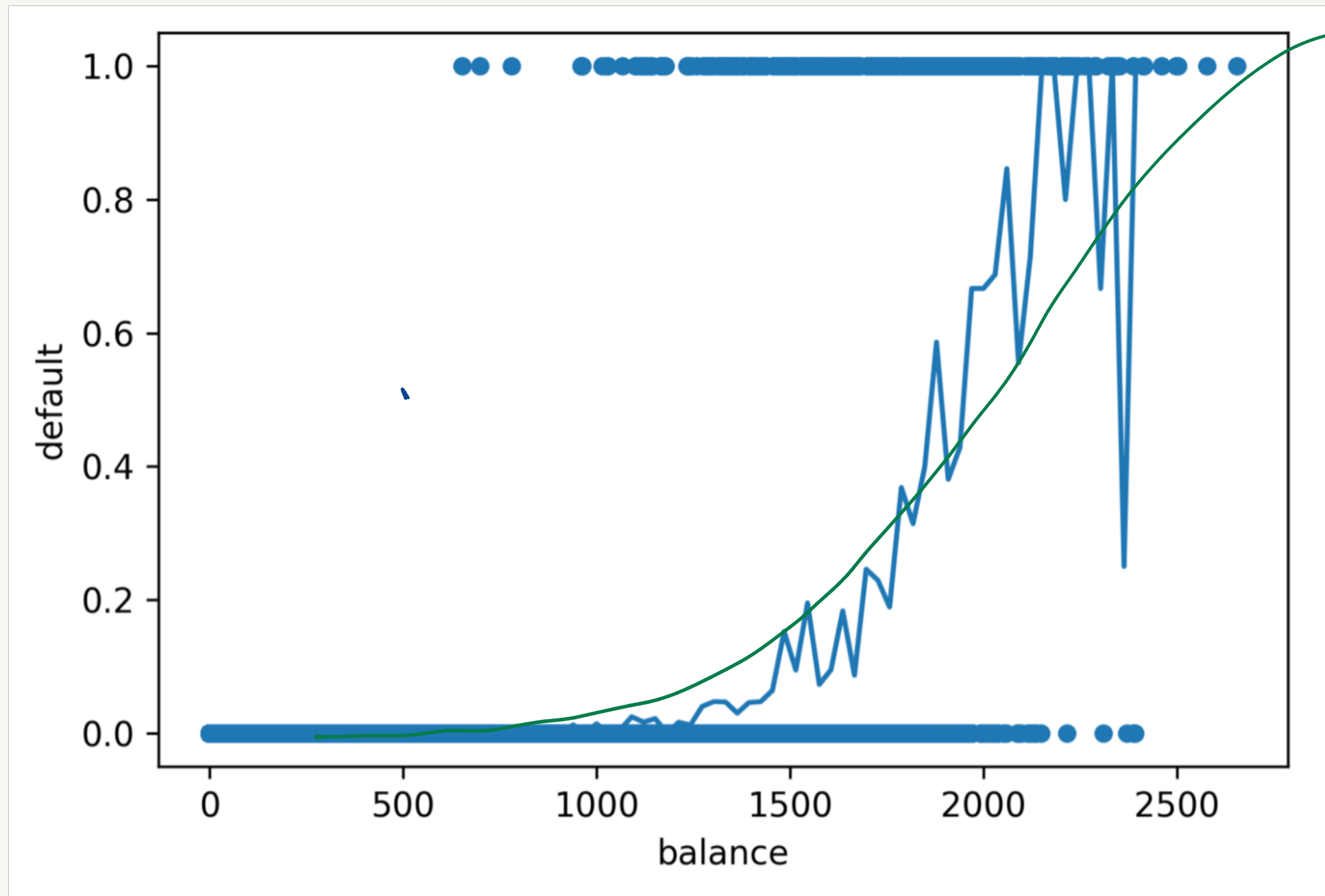
$$\beta_0 + \beta_1 x_1 = 0$$

$$\beta_1 x_1 = -\beta_0$$

$$\Rightarrow x_1 = \frac{-\beta_0}{\beta_1}$$

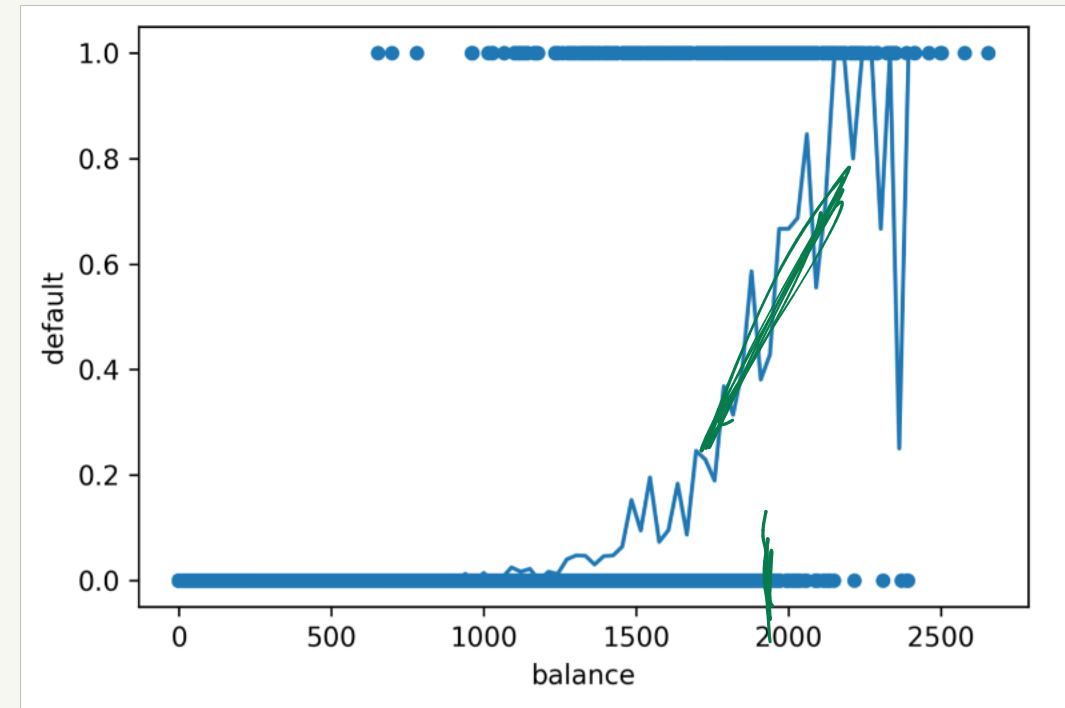
Fra forrige programmeringsøkt

- 1) Sigmoid
- 2) Berikningsgren.



# Manual regression

$$\text{def logiszt}(x, \beta_0, \beta_1):$$
$$\text{return } \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$



$$x\_vende = 1900$$

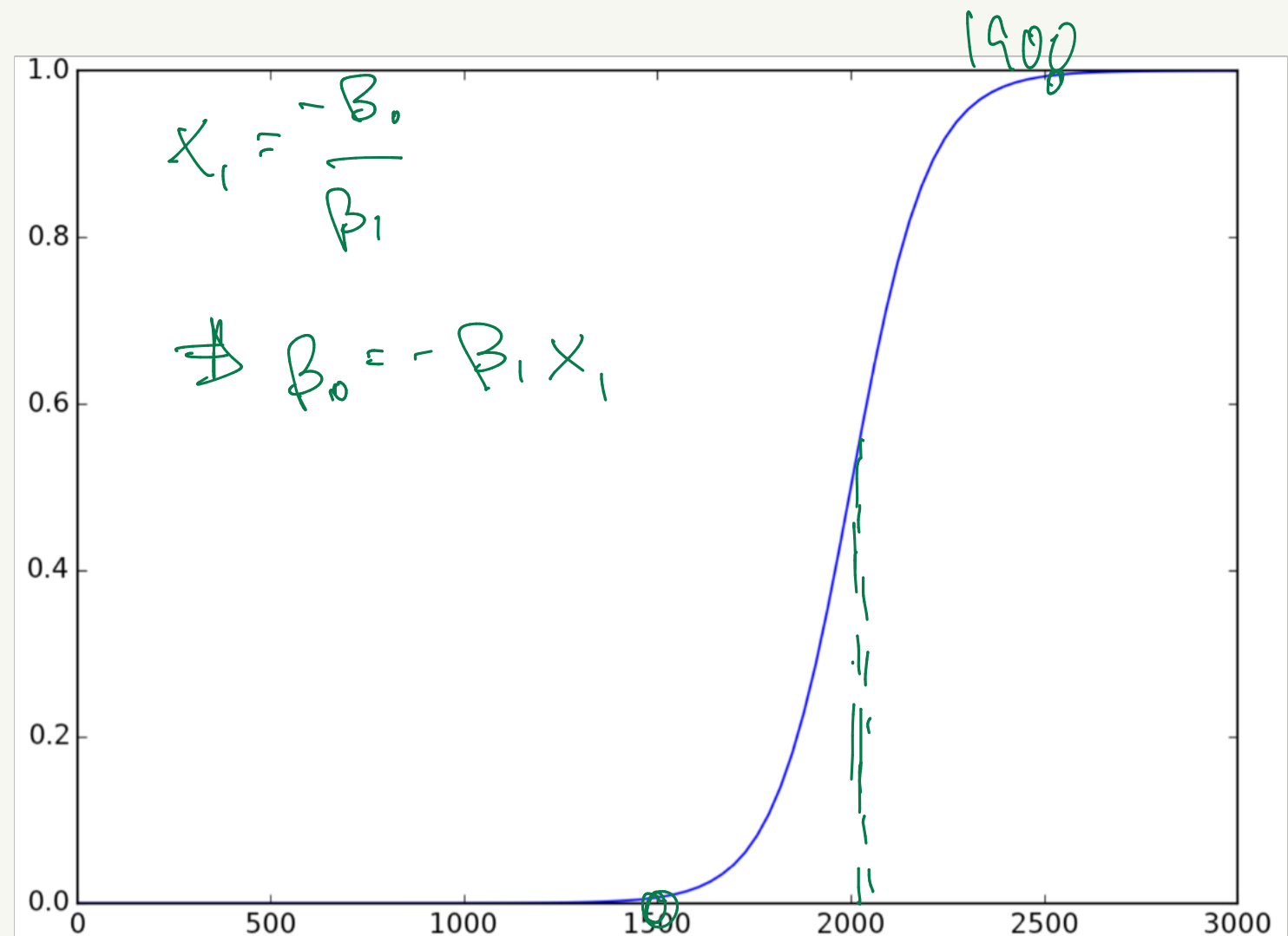
$$\beta_1 = 0,01$$

$$\beta_0 = \beta_1 \cdot x\_vende$$

$$x = \text{balance}$$

$$y = \text{logiszt}(x, \beta_0, \beta_1)$$

$$\text{plot}(x, y)$$



Live - programmering .

Logistisk regression

Følge med, ikke kode med -

# Eller vise løsninger

```
# antar vi her beregnet andel mislighold fra før
from sklearn.linear_model import LogisticRegression

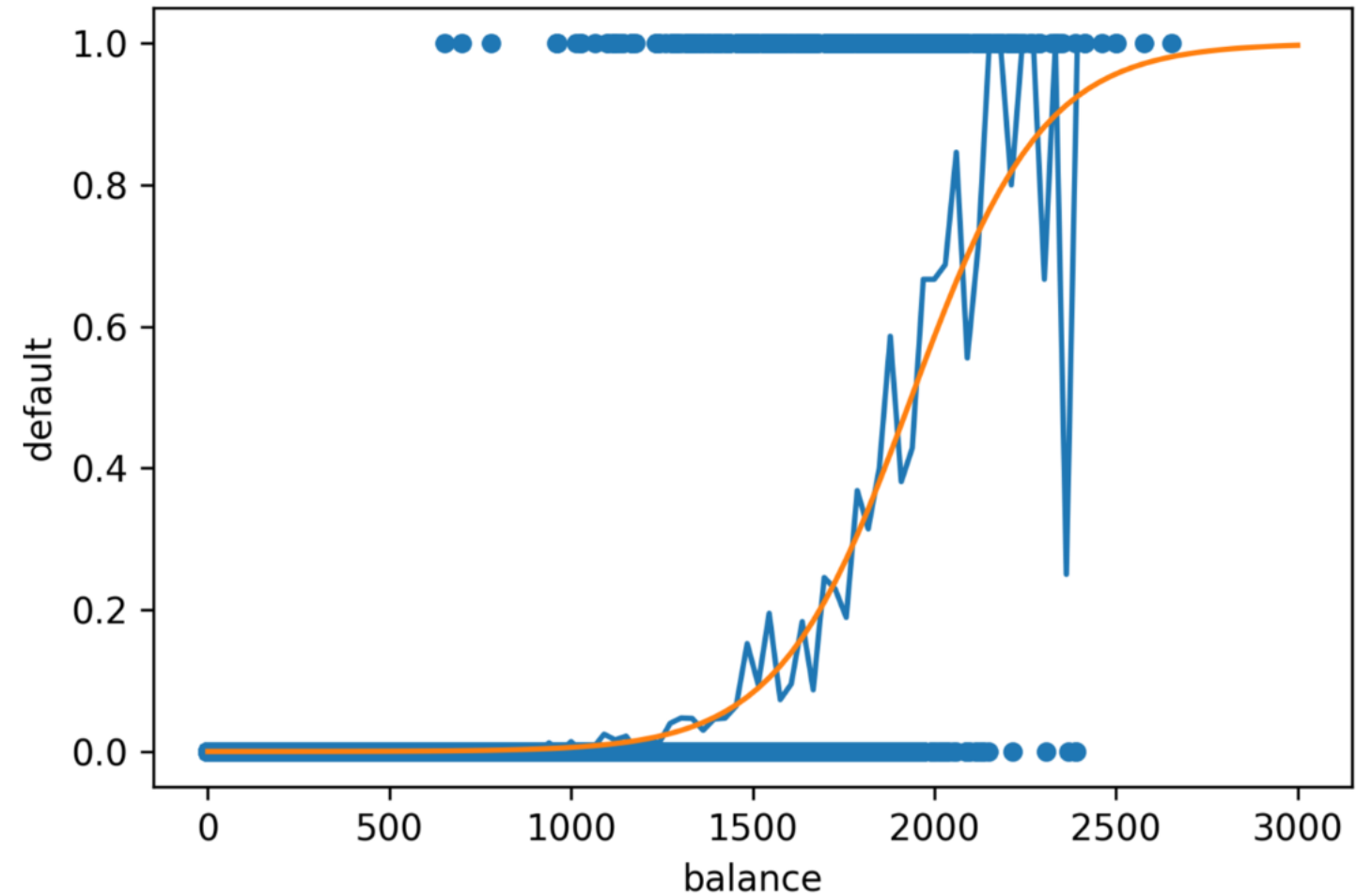
model = LogisticRegression()
fit = model.fit(data[["balance"]], data["default"])

print("Coefficients: ", model.coef_)
print("Intercept: ", model.intercept_)

beta0 = model.intercept_[0]
beta1 = model.coef_[0][0]

def sigmoid(x, beta0, beta1):
    z = beta0 + beta1 * x
    return np.exp(z) / (1 + np.exp(z))

data.plot.scatter("balance", "default")
plt.plot(x[:-1], andel_mislighold)
plt.plot(x, sigmoid(x, beta0, beta1))
plt.savefig("default.png", dpi=300)
```

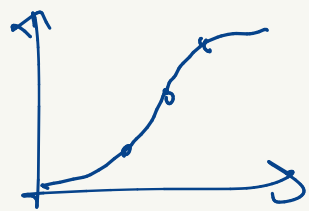




# To input - variable

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

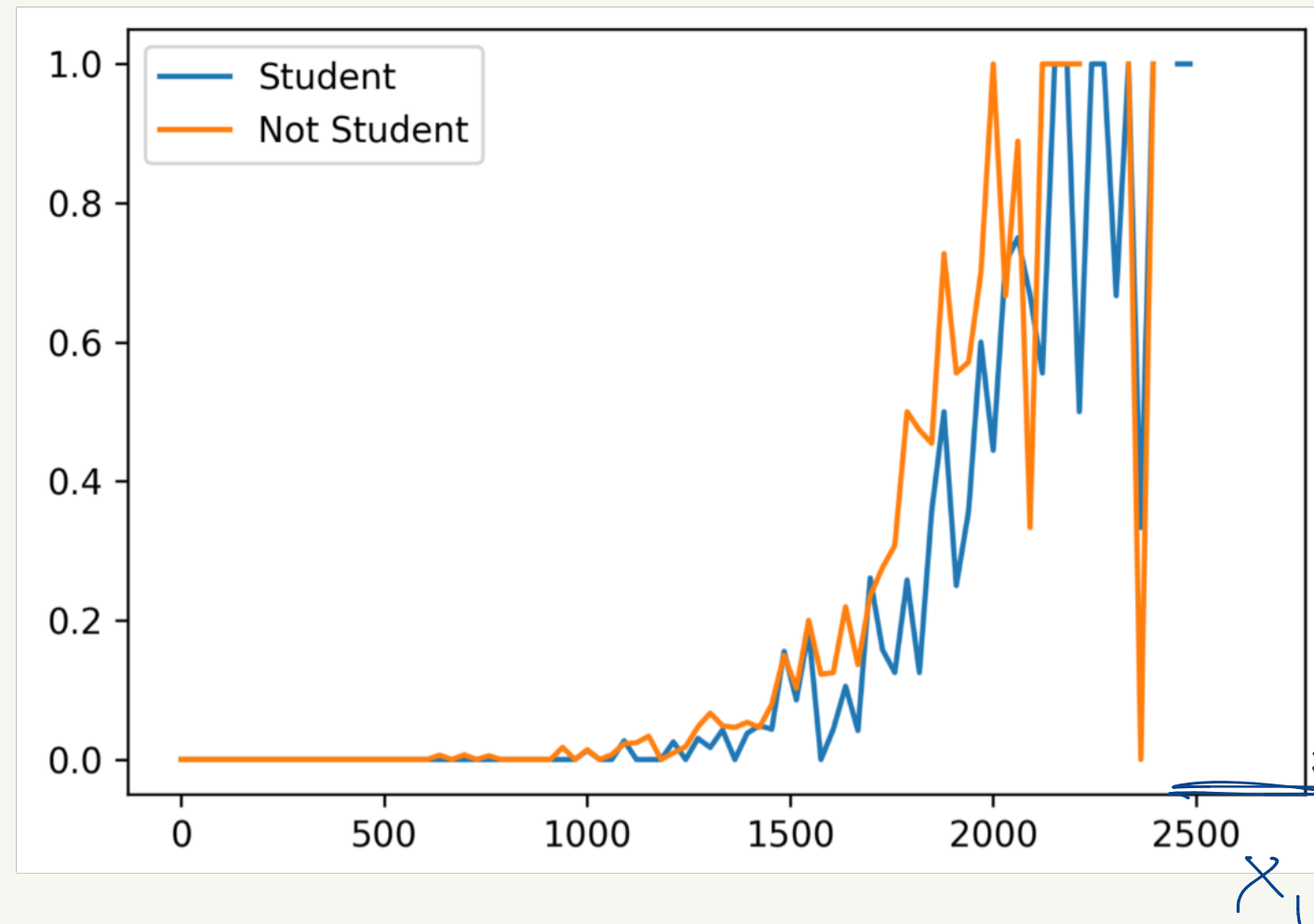
$$f(z) = \frac{e^z}{1 + e^z}$$



$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Intercept.      Sensitivity for saldo      Student?      Sensitivity for a rare student.

Saldo      Student?





# Underveis oppgave

- Er studenter, basert på dataene presentert her, bedre eller

dårligere betalere

enn andre folk?

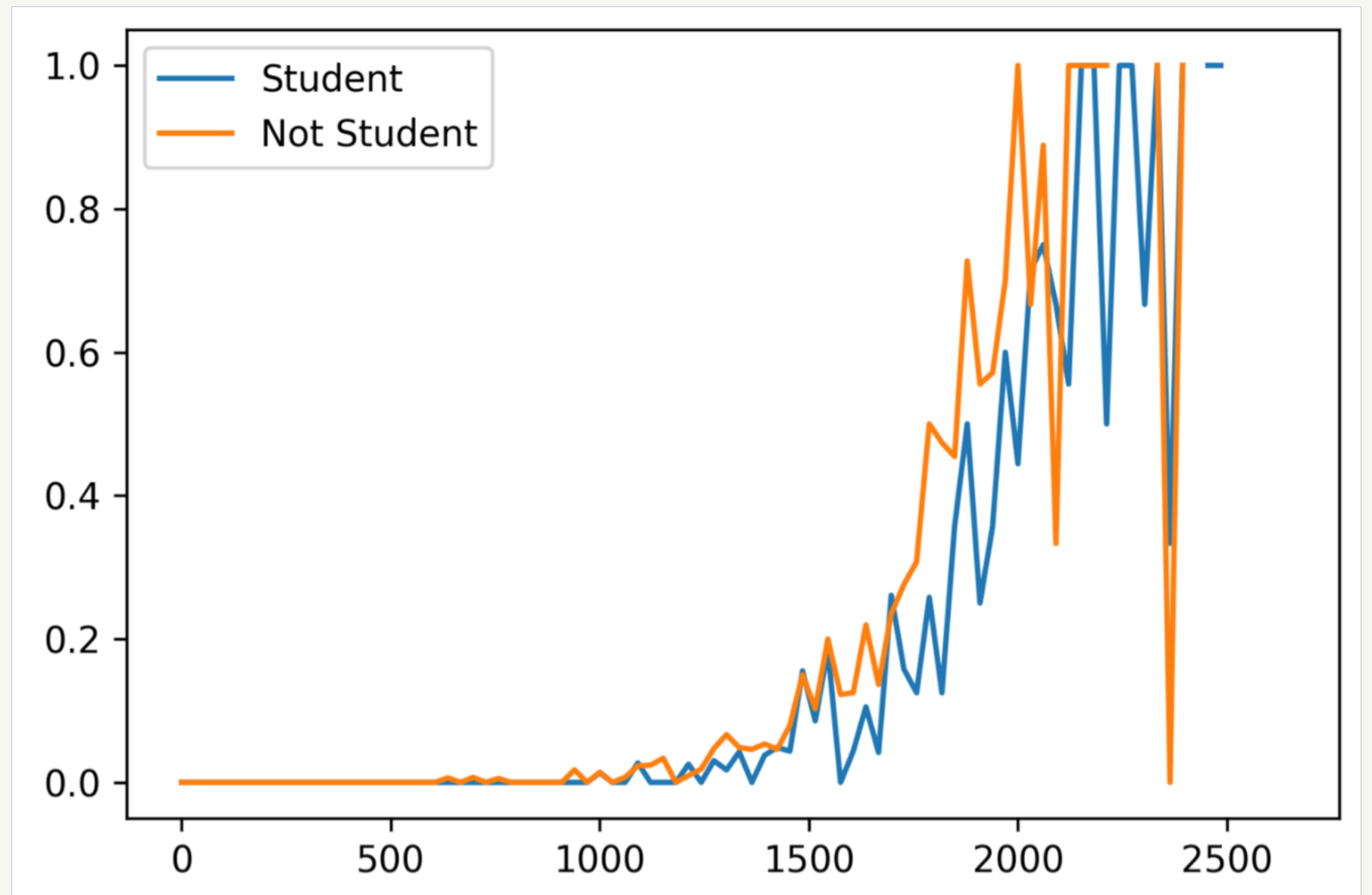
Tenke i 30 sekunder selv.



Bedre



Dårligere



# Crubleviseoppgave

Hva om  $X_1$  og  $X_2$  var kontinuerlige.

↑  
balanse

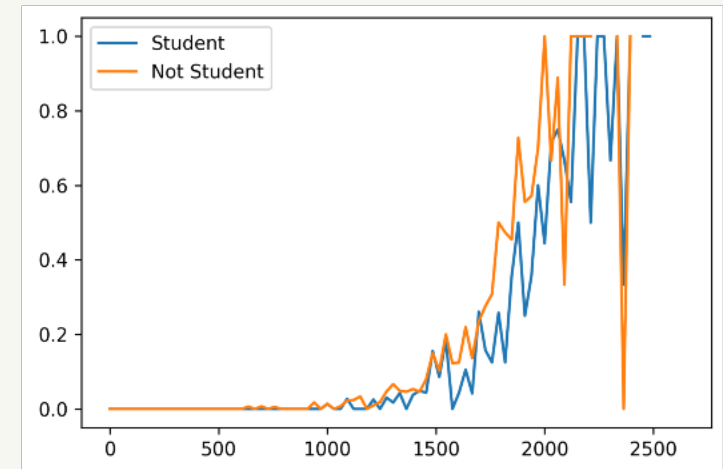
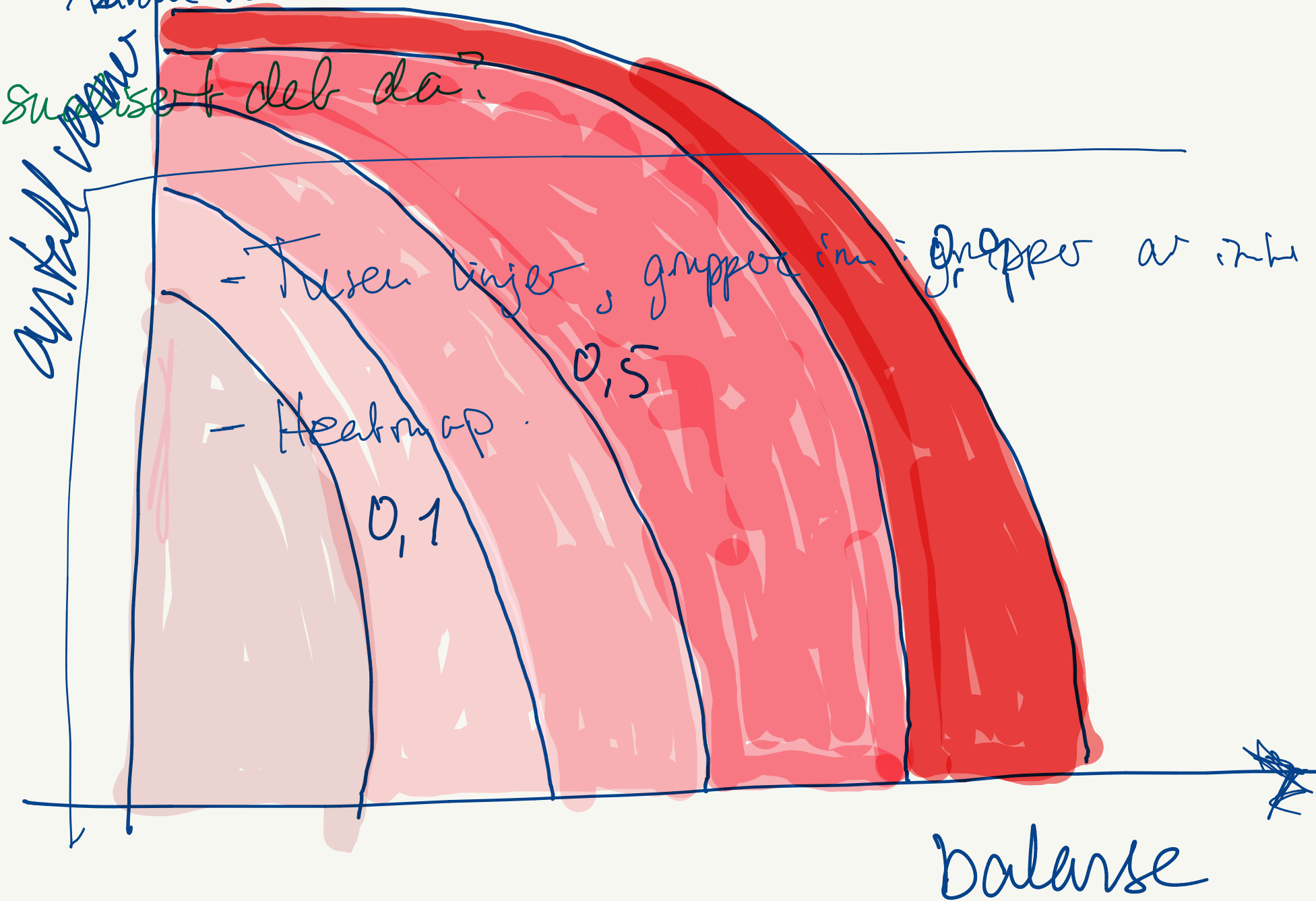
↑  
Antall venner..

Hvordan skulle vi visualisere det da?

1 min individuelt.

2 min gruppe.

Rask oppsummering.



Live - programming

logistic regression n/to variable.

# Tolkning av koefficientene

$$\beta_0 = -10.7$$

$$\beta_1 = 0.0057$$

$$\beta_2 = -0.699$$

$\sim 1000$

$\sim 5$

$\sim$

$\sim -0.7$

$$Z = \beta_0 + \underbrace{\beta_1 \cdot x_1}_{\text{Skala}} + \underbrace{\beta_2 \cdot x_2}_{[0,1]}$$

$$1000 \cdot \beta_1 \approx 5$$

$0 \vee 0.699$

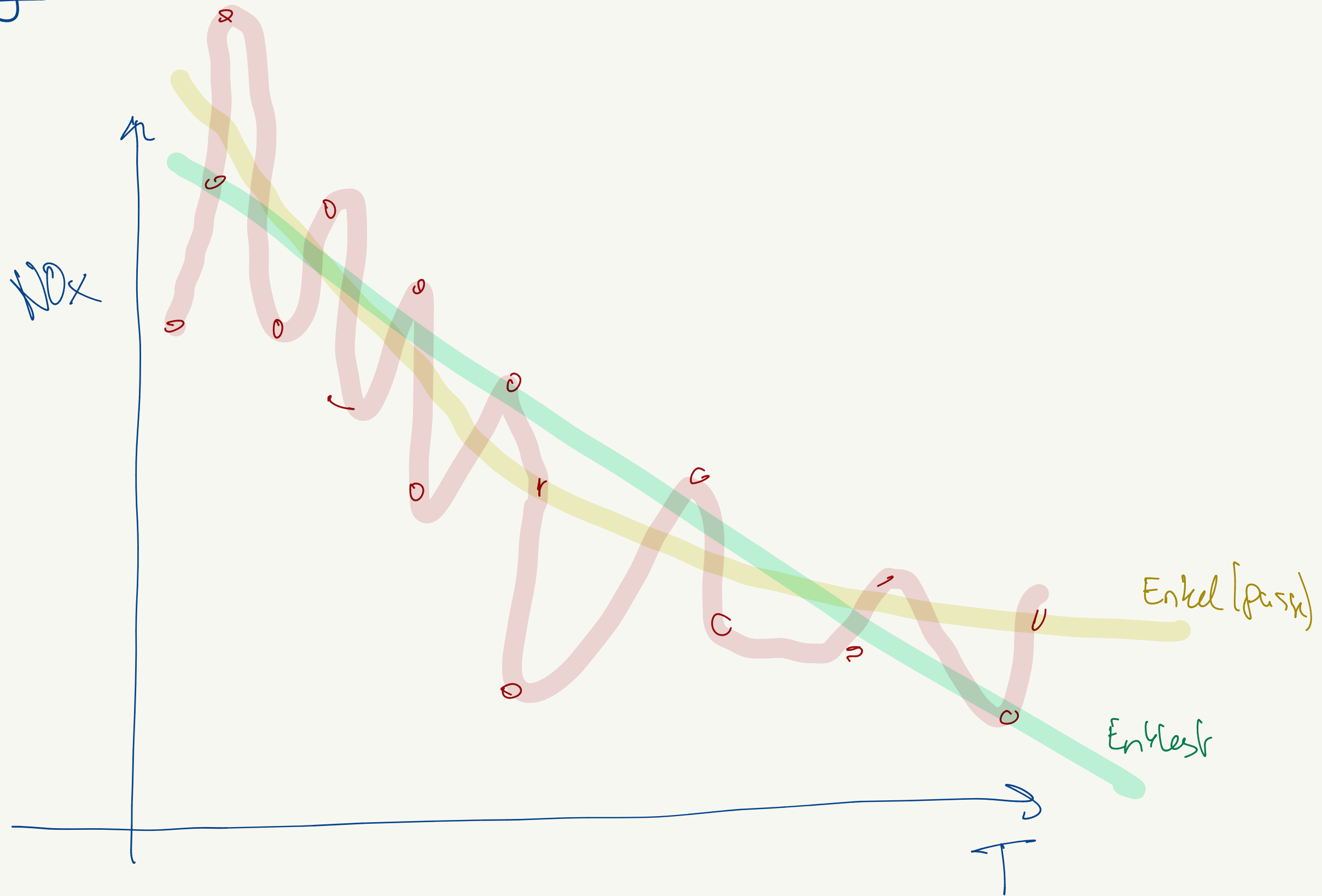
$\approx -0.7$

# Enda fleve variables

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots$$

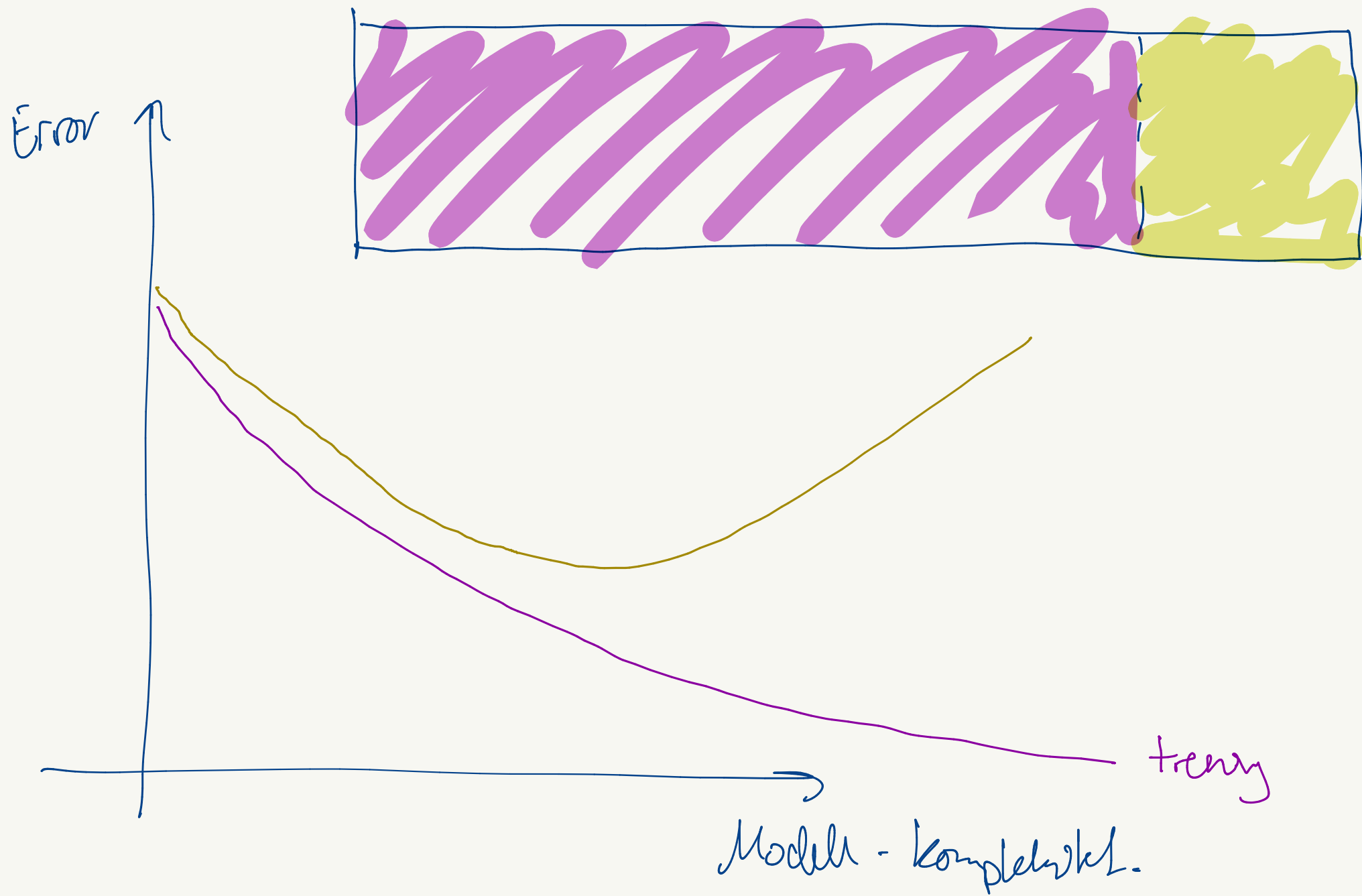
# Overfitting

$$y = f(x) + \varepsilon$$



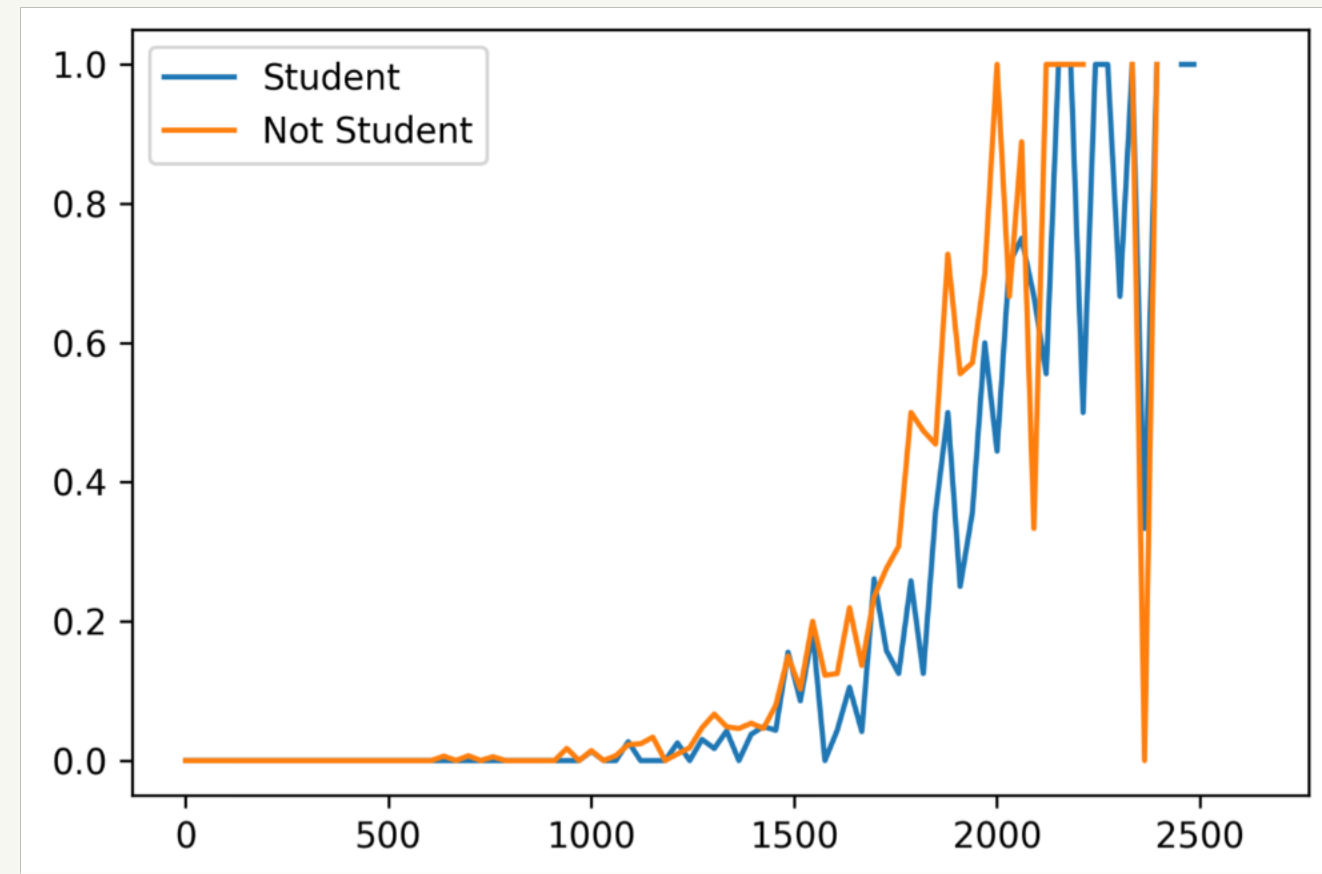


# Split



One hot encoding

# Beslutningsstr



Live - kode beslutningsstre.

# Oblig 1 - Titanic

Ordnal home:

Tuesday 14-16

Friday 10-12

# Prosjekt 1

Zip-fil

## Student karakterer

- Lage en modell med XGboost.

## Compas

- Behagel sannsynlighet
- Evaluer en eksisterende modell

Hva er det viktig å bruke modeller til?

Sende mail til Henrik med ønske

- Prosjekt

- Medarbeidere

- Studierching

- Fys-stk -3155/andre masterkurs?

- Eksamen

4 veiledningsøker

Muntlig eksamen  
13 april for prosjekt 1

Innleveringsfrist 11. april