3) $\underline{x} \in \mathbb{R}^N$ og $\underline{a} \in \mathbb{R}^N$

$$\underline{x} = [x_1, x_2, \ldots, x_N]$$

$$\underline{a} = [a_1, a_2, \ldots, a_N]$$

$$\underline{x}^T \cdot \underline{a} = x_1 a_1 + x_2 a_2 + \ldots + x_N a_N$$

$$\frac{\partial}{\partial a} = \left[\frac{\partial}{\partial a_1}, \frac{\partial}{\partial a_2}, \frac{\partial}{\partial a_3}, \ldots, \frac{\partial}{\partial a_N}\right]$$

$$\frac{\partial}{\partial a}(\underline{x}^T \underline{a}) = \left[\frac{\partial}{\partial a_1}(x_1 a_1 + \cancel{x_2 a_2 + \ldots}), \right.$$

$$\frac{\partial}{\partial a_2}(\cancel{x_1 a_1} + x_2 a_2 + \cancel{\ldots}),$$

$$\vdots$$

$$\left. \frac{\partial}{\partial a_N}(\cancel{x_1 a_1 + x_2 a_2 +} \ldots x_N a_N) \right]$$

$$= [x_1, x_2, \ldots, x_N] = \underline{x}$$

4) $\hat{P}(c_1|x) > \theta$, $0 < \theta < 1$ ②

$FP = N\cos\left(\frac{\pi\theta}{2}\right)$, N: # Negative samples

$TP = T\cos\left(\frac{\pi\theta}{2}\right)\left(2 - \cos\left(\frac{\pi\theta}{2}\right)\right)$, T : # Pos. samples

a) Def: $f_p = \frac{FP}{\underbrace{FP + TN}}$ og $t_p = \frac{TP}{\underbrace{TP+FN}}$

Konfusion matrix:

	TP	FP
	FN	TN

$N = TN + FP$

$T = TP + FN$

$\to f_p = \frac{\cancel{N}\cos\left(\frac{\pi\theta}{2}\right)}{\cancel{N}} = \cos\left(\frac{\pi\theta}{2}\right)$ x

$t_p = \frac{T\cos\left(\frac{\pi\theta}{2}\right)\left(2 - \cos\left(\frac{\pi\theta}{2}\right)\right)}{T}$

$= \underbrace{\cos\left(\frac{\pi\theta}{2}\right)}_{f_p}\left(2 - \underbrace{\cos\left(\frac{\pi\theta}{2}\right)}_{f_p}\right)$ y

b) ROC - curve : La θ variere mellom $[0,1]$

Plot t_p mot f_p.

c) AUC - score : $\int_0^1 y(x)\,dx$ (arealet under "kurven")

$$y = t_p = f_p(2 - f_p) = x(2 - x) = -x^2 + 2x$$

$$A = \int_0^1 (-x^2 + 2x)\,dx = \left[-\frac{1}{3}x^3 + \frac{1}{2}\cdot 2x\right]_0^1 = 1 - \frac{1}{3} = \underline{\underline{\frac{2}{3}}}$$

Soft - classifications - x. csv

	Ground truth		Probability
$x^{(1)}$	1	$\hat{p}^{(1)}$	0,9928
$x^{(2)}$	1	$\hat{p}^{(2)}$	0,99540
$x^{(3)}$	1	$\hat{p}^{(3)}$	0,97839
$x^{(4)}$	0	$\hat{p}^{(4)}$	0,53893
⋮			

Hvis terskel $\theta = 0,5$ så vil denne vær feil - klassifisert

Rett klassifisert om $\theta = 0,6$ derimot

True positives

False positives

Betyr at TP - rate og FP - rate avhenger av terskelverdi θ for klassifisering som True/False!

$$t_p = \frac{TP}{TP + FN}$$

Totalt # egentlige positive

Skulle egentlig vært positive!

$$f_p = \frac{FP}{FP + TN}$$

Tot # eg. negative

Skulle eg. vært negativ

t_p kalles også "Recall" i literaturen, eller også sensitivitet. Innen medisin er dette en faktor som man ønsker høyest mulig! (kan også støte på begrepet "precision" $= \frac{TP}{TP + FP}$)

Ofte fører økt recall til lavere presisjon og motsatt ...)

ROC står for Receiver Operatating Characteristic (5)

og plotter t_p mot f_p, altså t_p på y-akse og f_p på x-akse.

5a) Forslag algoritme:

1) Bruk første kollonne til å dele ssh. kollonne i samples $\underline{p}_+$ og $\underline{p}_\div$

2) True positives finner vi i subsettet $\underline{p}_+ > \theta$ og true negatives i $\underline{p}_\div \leq \theta$

Lengden av disse blir da TP og TN.

3) FP finner vi da for $\underline{p}_\div > \theta$ og FN for $\underline{p}_+ \leq \theta$