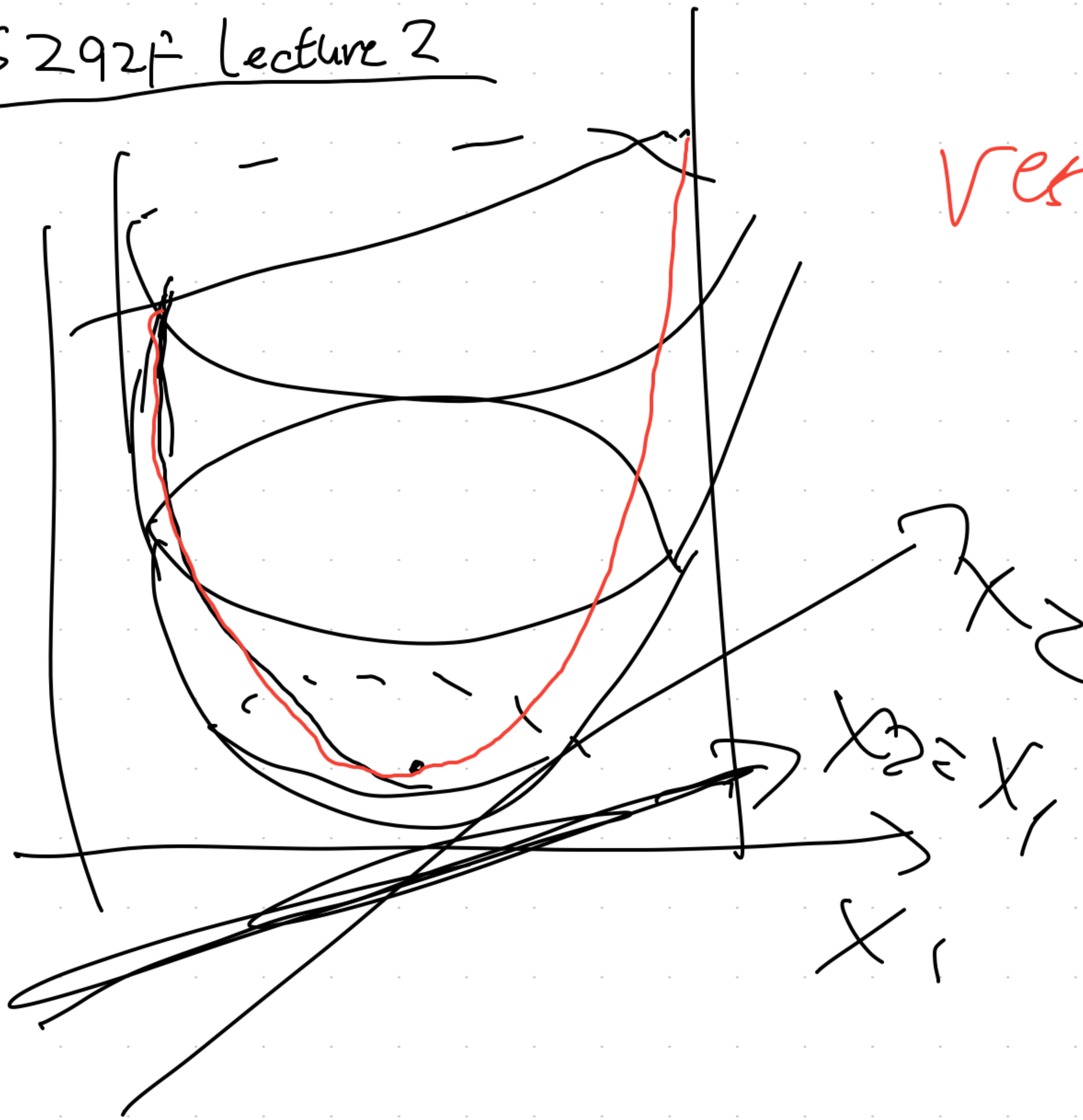
CS 292F Lecture 2



restriction to a line

$X_2 = X_1$

$h(g(x))$ $(h \circ g)'' \geq 0$ then $h \circ g$ is convex

$(h \circ g)'(x) = \underline{h'(g(x))} \cdot \underline{g'(x)}$ by chain rule

$$\underbrace{\underbrace{h''(g(x))}_{\geq 0}\,\underbrace{(g'(x))^2}_{\geq 0}} + \underbrace{\underbrace{h'(g(x))}_{\geq 0} \cdot \underbrace{g''(x)}_{\geq 0}} \geq \underline{0}$$

? when

h convex, h nondecreasing, g convex $\Rightarrow$ $h \circ g$ is convex

h convex, h nonincreasing, $\underline{g \text{ concave}}$

≥ 0 ≥ 0 $\leq 0 \cdot \leq 0$

≥ 0

$h \circ g$ is convex

$$f(x) = \log\left(\sum_{i=1}^{n} \exp(x_i)\right) \quad \text{“softmax”}$$

$$\approx \max_i \{x_i\} \leq$$

not diff.

$$\log(\exp(x_{\text{argmax}\{x_i\}})) \leq \log(\sum \exp(x_i)) \leq \log(n \cdot \max_i(\exp(x_i)))$$

$$\leq \underline{\log n} + \max_i(x_i)$$

$$\log(\exp(x_{\text{argmax}\{x_i\}})) = \max_i x_i$$

“soft-argmax” : $R^n \to R^n$ probability

$x = [1, 3, 2]$

$$\frac{e^{x_i}}{\sum_{i=1}^{n} e^{x_i}} = \begin{pmatrix} \frac{e^1}{e^1+e^3+e^2} \\ \frac{e^3}{e^1+e^3+e^2} \\ \frac{e^2}{e^1+e^3+e^2} \end{pmatrix} \approx \text{“argmax”}(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Softmax

Transform, in ML

First order optimality condition: diff. f, f convex

$$x \in \underset{x \in C}{\operatorname{argmin}} f(x) \iff \underline{\nabla f(x)^T (y-x) \geq 0} \text{ for any } y \in C$$

$$f \text{ convex} \Rightarrow f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

$$\geq f(x) \quad \forall y \in C$$

thus x is optimal

Special case: when $C = R^n$

$$\nabla f(x)^T (y-x) \geq 0 \iff \nabla f(x)^T a \geq 0 \quad \underline{\forall a \in R^n}$$

$$\iff \nabla f(x)^T a = 0 \qquad \underline{y - x} = a$$

$\forall a \in R^n$ $\quad R^n$

$a, -a$

$$\iff \nabla f(x) = 0$$

$\square$

By FOC, x satisfy $\nabla f(x)^T(y-x) \geq 0 \quad \forall y$ s.t. $Ay=b$ $\qquad x = A^+ b + v$, $v \in null(A)$

$\Longleftrightarrow$

$\nabla f(x)^T v = 0 \quad \forall v \in null(A)$ $\qquad Av = 0$

$y - x = \cancel{A^+ b} + v_y - (\cancel{A^+ b} + v_x) = \underbrace{v_y - v_x}_{\in null(A)}$

$\nabla f(x)$

$\nabla f(x) \in null(A)^\perp = row(A) = \underline{col(A^T)}$

$\exists$ coefficient $\tilde{u}$

s.t. $\nabla f(x) = A^T \tilde{u}$

$$\boxed{\nabla f(x) + A^T u = 0}$$

method of Lagrange Multiplier

$u = -\tilde{u}$