Equivalent definitions of convexity

Showing a function is convex

Gradient Methods

"Strictly" convex

$f(\theta x+(1-\theta)y) < \theta f(x) +$

-implies at most one g

"Concave": $(-f)$ is co

"Log-convex": $(\log f(x))$

- Pick up A2 to mark
- A3 due now.
- Pick A3 and course eval at end of class. (return now or Monday)
- No tutorials this week (A4 out tomorrow, due Oct 15)
- Why do support vectors make kernel methods more effici

Usual

$$k(X,X) = \begin{bmatrix} & & N \\ & & \\ & & \\ N & & \end{bmatrix}$$

Support vectors

$$k(X_{sv},X_{sv}) = \begin{bmatrix} & & \#sv \\ & & \\ & & \\ \#sv & & \end{bmatrix}$$

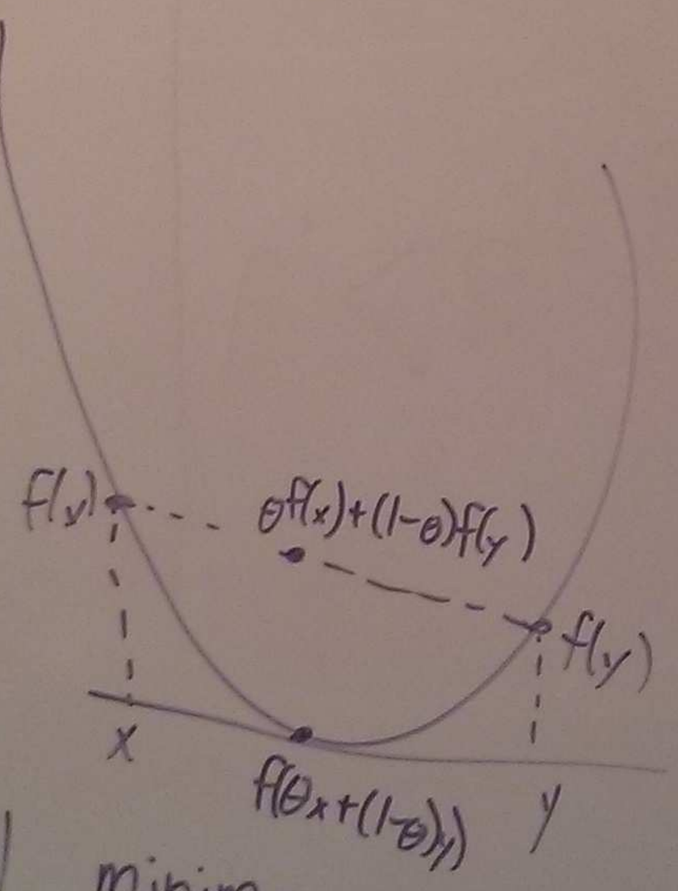
$$k(\hat{X},X_{sv}) = \begin{bmatrix} & & \#sv \\ & & \\ T & & \end{bmatrix}$$

Convex Functions (zero-order definition)

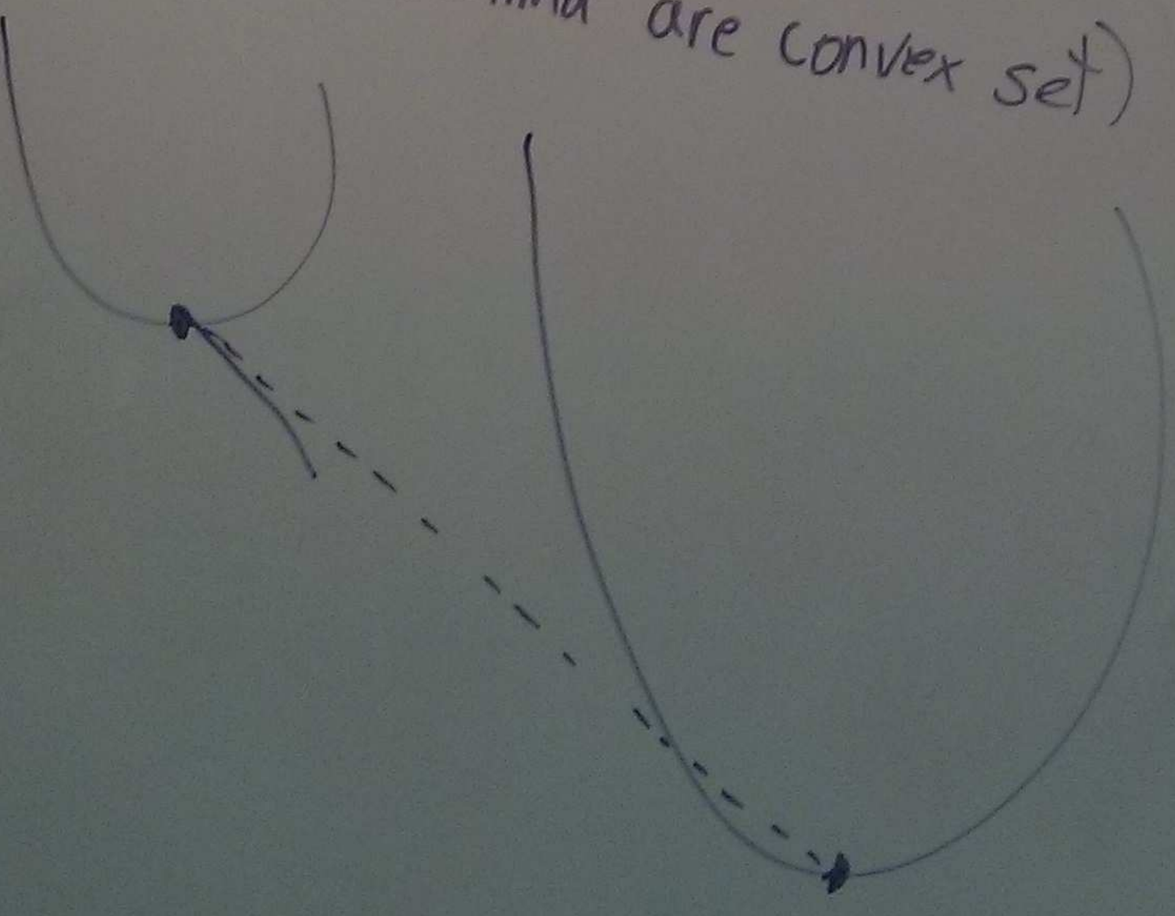
Function 'f' is convex if dom(f) is convex, and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

"below the chord", "epigraph is convex set"



- Implies local minima are global minima
(and global minima are convex set)



Examples:

$f(x) = \exp(ax)$

$f(x) = x \log x, \; x > 0$

$f(x) = ax^2 + bx, \; a \geq 0$

$f(\bar{x}) = \max_i \{x_i\}$

$f(\bar{x}) = \log\left(\sum_{i=1}^{d} \exp(x_i)\right)$

$f(x) = -\log(x)$

"Strictly" convex

$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$.

- implies at most one global minimum

"Concave": $(-f)$ is convex

"Log-convex": $(\log f(x))$ is convex

Co

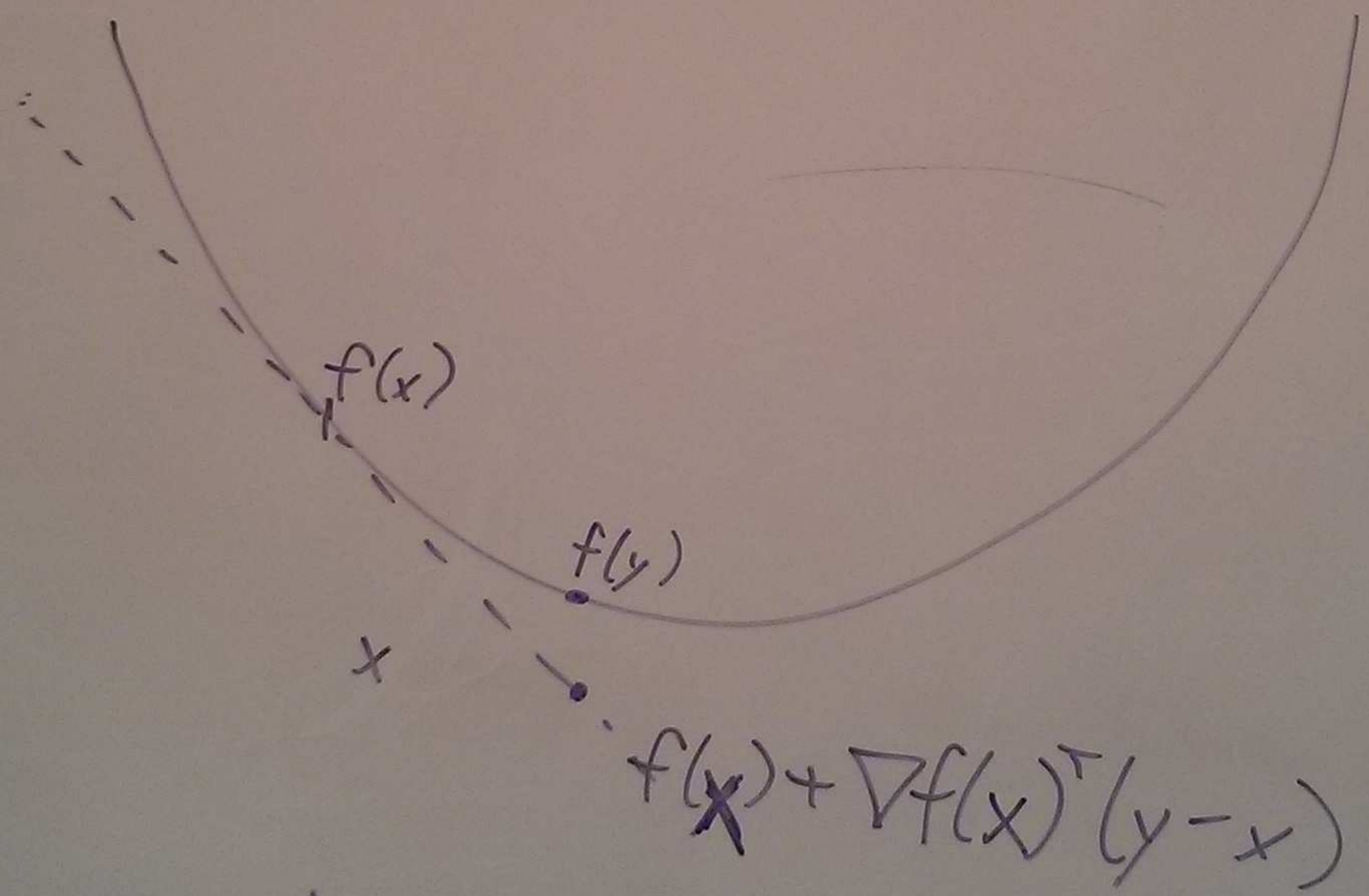
Fu

Convex Functions (first-order definition)

If 'f' is differentiable, equivalent to

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in \text{dom}(f)$$

"f is above its tangent"



- Can see that any stationary point ($\nabla f(x^*) = 0$) is a global optimum,

$$f(y) \geq f(x^*) + 0, \quad \forall y$$

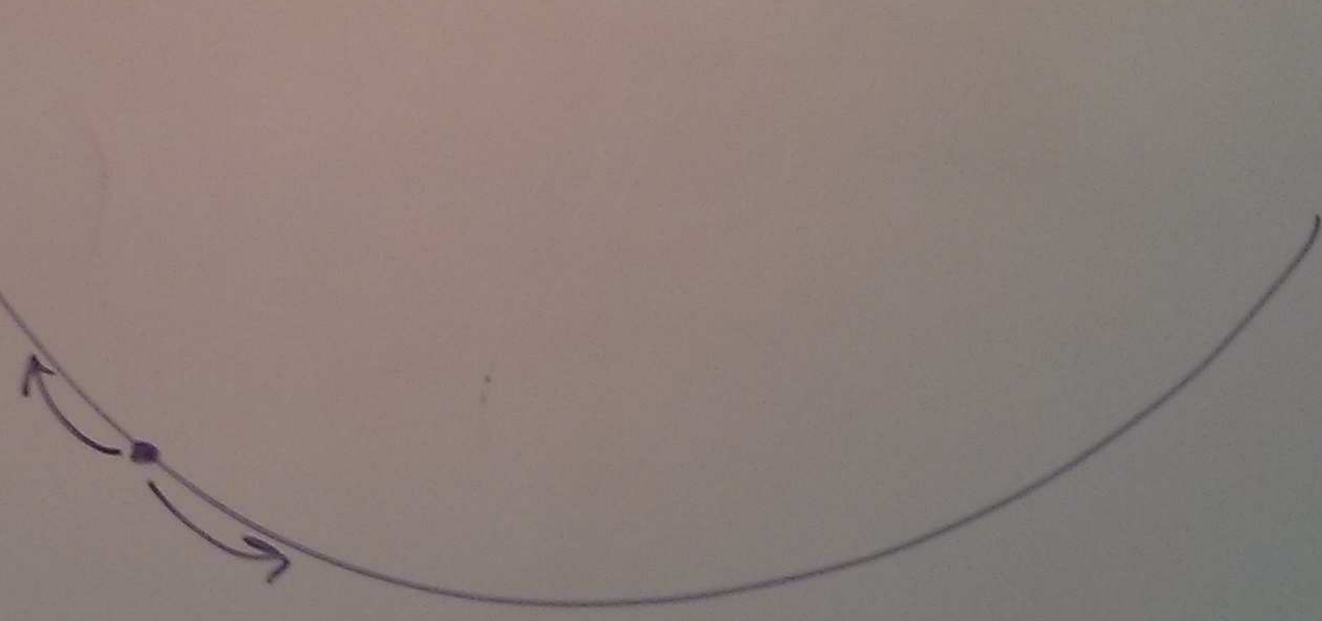
Conv

Convex Functions (second order definition)

If 'f' is twice-differentiable, equivalent to

$$\nabla^2 f(x) \geq 0, \forall x \qquad (f''(x) \geq 0, \text{ fo}$$

"f is curved upward everywhere"

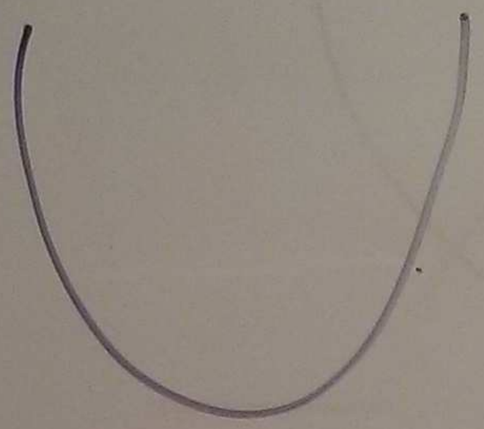


0) - usually the easiest way to show a simple function is convex

a function is convex

ient Methods

$f(x) = x^2$



$f'(x) = 2x$

$f''(x) = 2 \geq 0$

$f(x) = \frac{1}{2} x^T A x + b^T x + c, \; A \succeq 0$

$\nabla f(x) = Ax + b$

$\nabla^2 f(x) = A \succeq 0$

$$f(x) = \frac{1}{2}\|Ax - b\|^2$$

$$\nabla f(x) = A^T A x - A^T b$$

$$\nabla^2 f(x) = \underline{A^T A}$$

$$\overset{*}{=} \sum_{i=1}^{N} a_i a_i^T$$

$$x^T \left(\sum_{i=1}^{N} a_i a_i^T \right) x$$

$$= \sum_{i=1}^{N} (x^T a_i)(a_i^T x) \geq 0$$

$I_v =$

$x^T A x \geq 0, \forall x$

Operations that preserve convexity

Let f_1, f_2 be convex functions

1. Non-negative weighted sum: $w_1 f_1(x) + w_2 f_2(x)$, $w_1, w_2 \geq 0$
2. Composition w/ Affine function: $f_1(Ax + b)$
3. Pointwise maximum: $\max\{f_1(x), f_2(x)\}$
4. Other composition rules.

Show that SVMs are convex:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \max\{0, 1 - y_i \bar{w}^T \bar{x}_i\}$$

① $w^T w = w^T I w$

$\nabla^2[w^T w] = 2I$

① $\max\{0, 1 - y_i \bar{w}^T \bar{x}_i\}$

③

$w^T A w$

0

$1 - y_i \bar{w}^T \bar{x}_i$

Gradient Method

- We want to minimize convex 'f'
(e.g., MLE, MAP, SVMs, etc.)

- Generate a sequence

$x^0, x^1, x^2, x^3, \ldots \rightarrow x^*$ (optimal value)

such that $x^t \rightarrow x^*$ as $t \rightarrow \infty$

- Gradient method:

$$x^{t+1} = x^t - \alpha^t \nabla f(x^t)$$

"step size" (small)

$x^* = \arg\min_x f(x)$

...ctions of convexity

...unction is convex

...Methods

Stochastic Gradient Method

$$\underset{x}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{N} f_i(x) = f(x)$$

$$x^{t+1} = x^t - \alpha \nabla f_i(x^t)$$

picked randomly

Is this reasonable? from $i \in \{1, 2, \ldots, N\}$

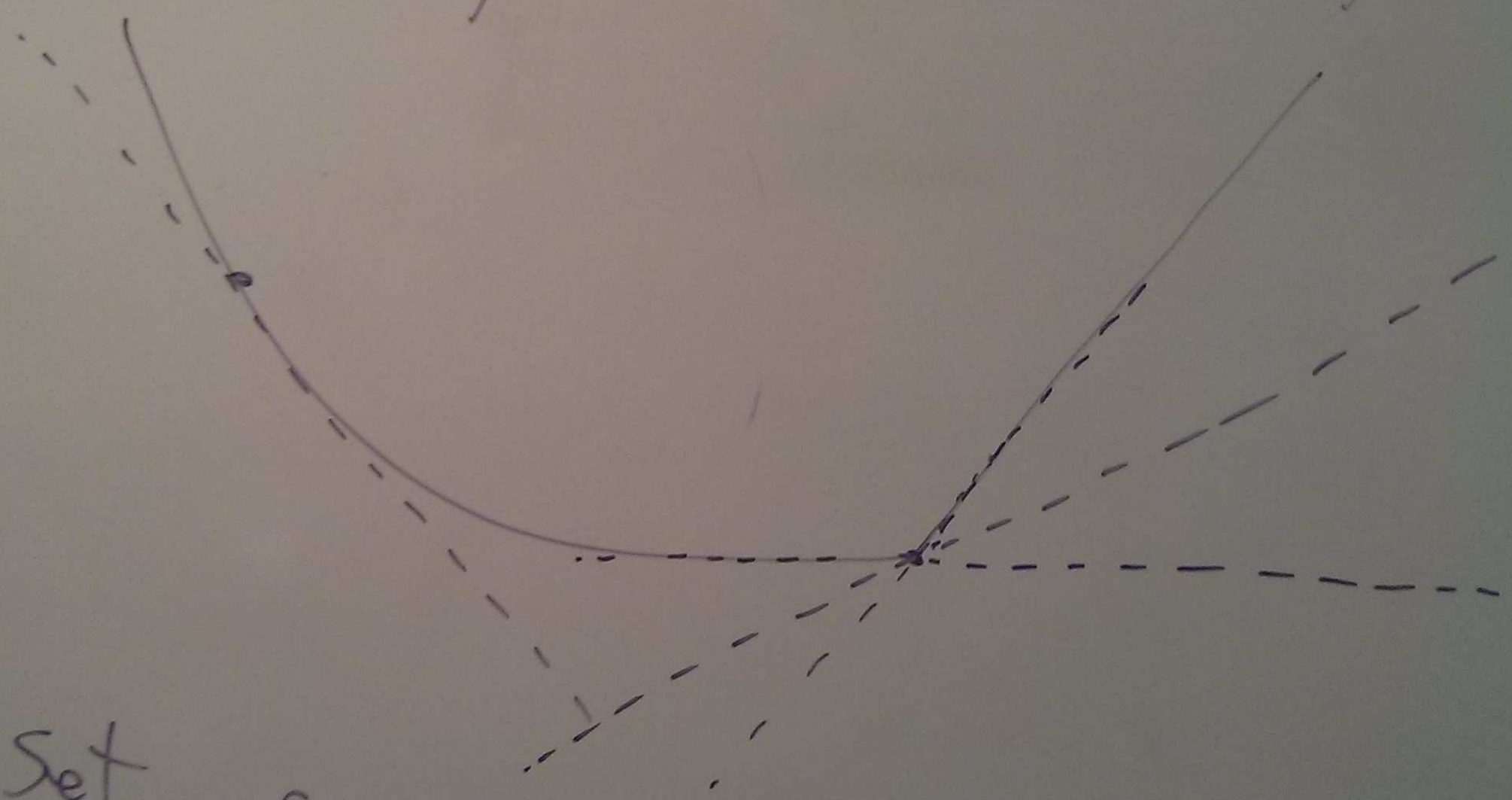
$$E\left[\nabla f_i(x^t)\right] = \sum_{i=1}^{N} p_i \nabla f_i(x^t)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \nabla f_i(x^t)$$

Subgradient

- a 'subgradient' of 'f' at 'x'
  is a vector 'd' such that

$$f(y) \geq f(x) + d^T(y - x$$



- Set of subgradients is the "sub-differe

$$\partial f(x) = \begin{cases} \{1\} & x > 0 \\ \{-1\} & x < 0 \\ [-1, 1] & x = 0 \end{cases}$$

x^* is Global optimum
iff $0 \in \partial f(x^*)$