Least squares

$$\underset{\vec{w}}{\text{minimize}} \quad \|\bar{X}\vec{w} - \vec{y}\|^2$$

$\boxed{\text{solution:}}$
- projection
- complete the square
- Δw argument

Quadratics: (for Ridge, weighted least squares)

$$\underset{\vec{x}}{\text{minimize}} \quad \tfrac{1}{2}\vec{x}^T \bar{Q} \vec{x} - \vec{p}^T \vec{x}$$

$\boxed{\text{solution:}}$
- complete the square
- Δw

PCA:

$$\text{maximize} \quad \|\bar{X}\vec{w}\|^2$$
$$\text{s.t.} \quad \|\vec{w}\|^2 = 1$$

$\boxed{\text{solution:}}$ Rayleigh Quotient

What about general problems? How do we

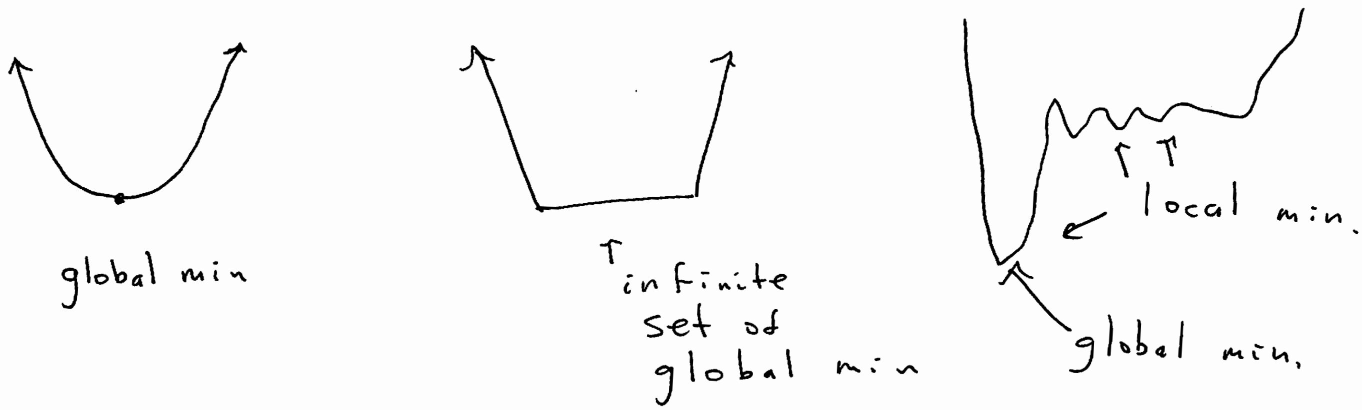
$$\underset{\vec{w}}{\text{minimize}} \quad f(\vec{w}) \quad ?$$

Optimization:

$$\underset{\vec{w}}{\text{minimize}} \quad f(\vec{w})$$

- $\vec{w}$ is a minimizer if $f(\vec{w}_\star) \le f(\vec{w}) \quad \forall w.$
- $\vec{w}$ is a local minimizer if, for some $R \geq 0$,

$$f(\vec{w}_\star) \le f(\vec{w}) \quad \forall \; w \; \text{s.t.} \; \|w - w_\star\| \le R$$

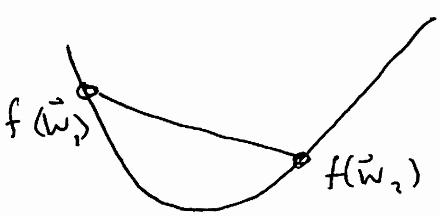


Note: 1-D pictures often misleading:

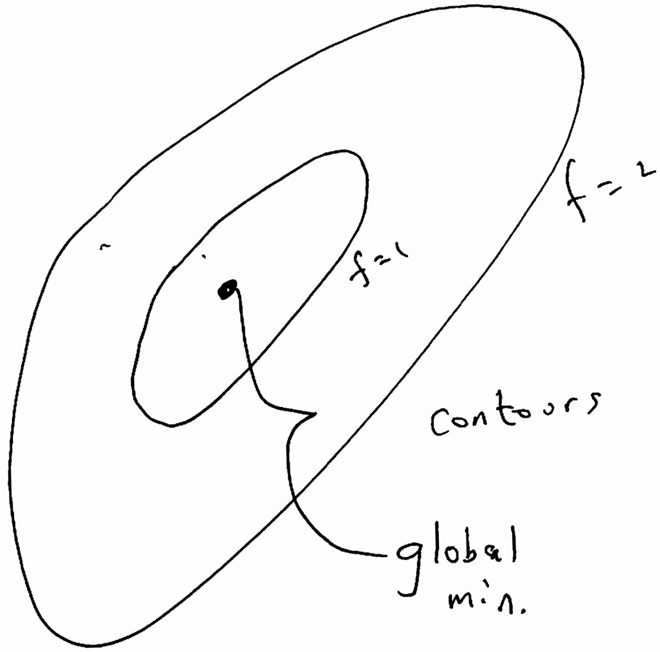
$f(w_1, w_2) = w_1^2$ has infinite number of global minimizers

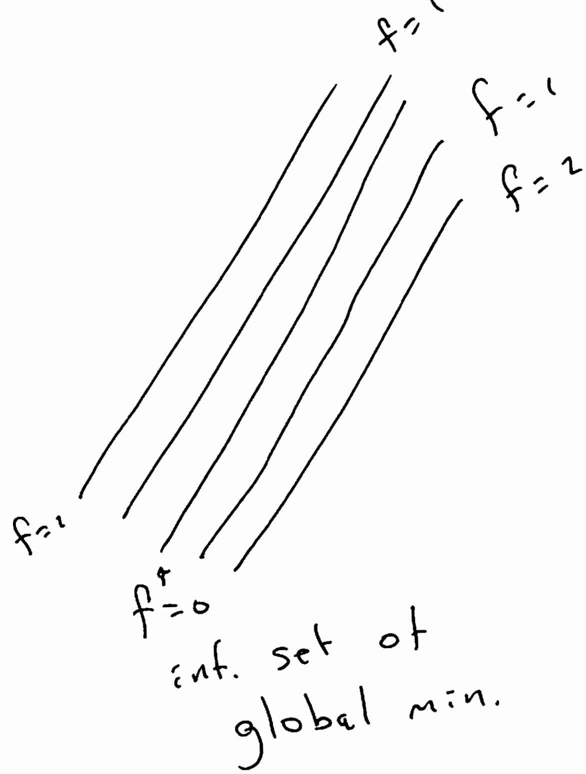
f is ~~conv~~ convex if $\forall \; \vec{w}_1, \vec{w}_2$ and $t \in [0,1]$

$$f(t\vec{w}_1 + (1-t)\vec{w}_2) \le t f(w_1) + (1-t) f(w_2)$$



line segment lies above the graph

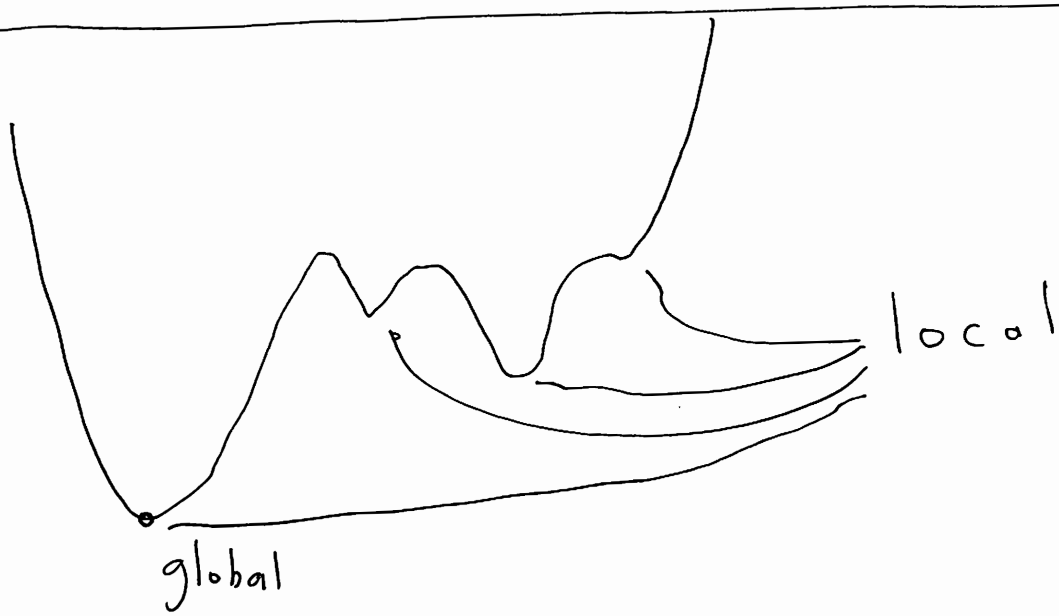




$$f(w_1, w_2) = (w_1 - w_2)^2$$

$$\underset{\vec{w}}{\text{minimize}} \; \|X\vec{w} - \vec{y}\|^2$$

$$n < d \implies \text{inf. global.}$$



Taylor's Theorem: Suppose f is twice continuously differentiable. Then $\forall\ w, w_0$, $\exists\ t \in (0,1)$ s.t.

$$f(w) = f(w_0) + \nabla f(w_0)^T (w - w_0) + \frac{1}{2}(w - w_0)^T \nabla^2 f(tw + (1-t)w_0)(w - w_0)$$

Heuristically, if w is close to w_0

$$f(w) \approx f(w_0) + \nabla f(w_0)^T (w - w_0) + o(\|w - w_0\|^2)$$

Minimization principle: w_* is a minimizer $\Rightarrow$

$\nabla f(w_*) = 0$

To see this: $\forall\ \Delta w$, $f(w_* + \Delta w) \geq f(w_*)$

By Taylor's Theorem this means

$$\nabla f(w_*)^T \Delta w + O(\|w\|^2) \geq 0$$

$$\Rightarrow \quad \nabla f(w_*)^T \frac{\Delta w}{\|w\|} + o(\|w\|) \geq 0$$

Since this holds for all Δw, letting Δw be arbitrarily small implies $\nabla f(w_*) = 0$

Note: Converse is not true! $f(w) = -w^2$

$\nabla f(0) = 0$, but 0 is a maximizer.

FACT: If f is convex, $\vec{w}_*$ is a minimizer iff $\nabla f(\vec{w}_*) = 0$.

PROOF: $t \in [0,1]$, $\vec{w}$ arbitrary.

$$f(\vec{w}_* + t(\vec{w} - \vec{w}_*)) = f((1-t)\vec{w}_* + t\vec{w})$$

$$\leq (1-t) f(\vec{w}_*) + t f(\vec{w})$$

$$\Rightarrow f(\vec{w}) \geq f(\vec{w}_*) + \frac{f(\vec{w}_* + t(w - w_*)) - f(w_*)}{t}$$

Taking the limit as $t \to 0$ yields

$$f(\vec{w}) \geq f(\vec{w}_*) + \underbrace{\nabla f(w_*)^T (w - w_*)}_{0}$$

///

$-\nabla f(w)$ always is a <u>descent</u> <u>direction</u>:

$$f(\vec{w} - t\nabla f(\vec{w})) = f(\vec{w}) - t\|\nabla f(\vec{w})\|^2 + t^2 o(\|\nabla f(\vec{w})\|^2)$$

For t small enough $f(\vec{w} - t\nabla f(\vec{w})) \leq f(\vec{w})$ as long as $\nabla f(\vec{w}) \neq 0$.

<u>ALGORITHM</u>: GRADIENT DESCENT

$$\vec{w}_{k+1} = \vec{w}_k - \alpha \nabla f(\vec{w}_k) \qquad \alpha : \text{step-size}$$

This either decreases the cost for some α or finds $\vec{w}$ with $\nabla f(\vec{w}) = 0$.

$$\underline{\underline{\vec{v}}}^T \Delta\vec{w} + o(\|\Delta\vec{w}\|^2) \geq 0$$

Least Squares

$$f(\vec{w}) = \|\bar{X}\vec{w} - \vec{y}\|^2$$

$$f(\vec{w}+\Delta\vec{w}) = \|\bar{X}(\vec{w}+\Delta\vec{w}) - \vec{y}\|^2$$

$$= f(\vec{w}) + \underline{2(\bar{X}^T(X\vec{w}-y))^T}\,\Delta\vec{w}$$

$$+ \underline{\|\bar{X}\Delta w\|^2}$$

$$\nabla f(\vec{w}) = 2\bar{X}^T(\bar{X}\vec{w} - \vec{y})$$

$$\nabla f(\vec{w}) = 0 \iff \bar{X}^T(\bar{X}\vec{w} - \vec{y}) = 0$$

normal equations

Quadratics:

$$f(\vec{w}) = \frac{1}{2}\vec{w}^T Q \vec{w} - \vec{p}^T \vec{w}$$

$$\nabla f(\vec{w}) = \bar{Q}\vec{w} - \vec{p}$$

$$\nabla f(\vec{w}) = 0 \iff \vec{w} = \bar{Q}^{-1}\vec{p}$$

Consider $h(\vec{w})$, suppose I know

$$\underbrace{h(\vec{w}) \le h(\vec{w}_0) + \nabla h(\vec{w}_0)^T (w - w_0) + \frac{L}{2}\|w - w_0\|^2}_{h_{up}(\vec{w})}$$

$$\nabla h_{up}(\vec{w}) = \nabla h(\vec{w}_0) + L(w - w_0)$$

$$\nabla h_{up}(\vec{w}) = 0 \iff \vec{w} = w_0 - \frac{1}{L}\nabla h(\vec{w}_0)$$