Week 5 - Lecture 23

A convex function is a function, that at any given point can give a lower bound. This lower boun is given by the tangent at any given point and is defined by the sub-gradients at that point.

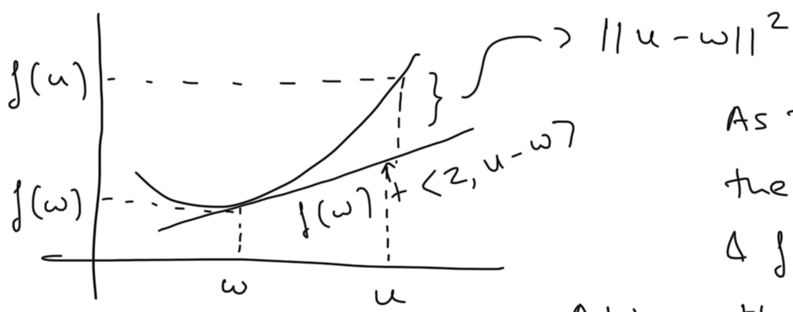
Strongly convex Functions

Defn :- A function $f : S \to R$ is σ - strongly convex ~~of~~ over S with respect to norm $\|\cdot\|$ if

$\forall\ w \in S$ we have

$\forall\ z \in \partial f(w)\ \ \forall\ u \in S$

$$\underbrace{f(u) \geq f(w) + \langle z, u-w \rangle}_{\text{This was already defined for convex function}} + \underbrace{\frac{\sigma}{2} \|u-w\|^2}_{\text{Increases the lower bound, still function is convex}}$$



As the gap increases, the diff. between $f(u)$ & $f(w)$ increases

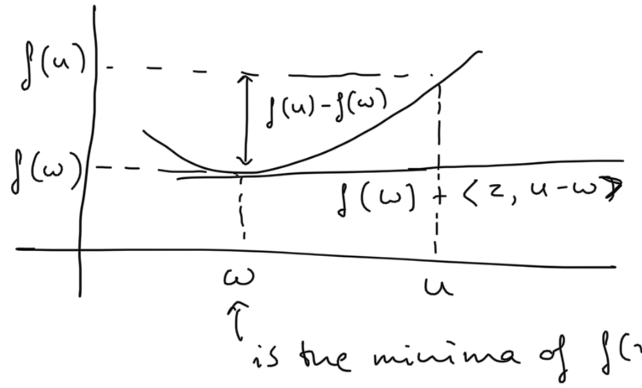
Apparently makes it easier to identify these functions.

Some properties of Strongly convex functions

Lemma: Let $\omega = \underset{x \in S}{\operatorname{argmin}} f(x)$. Then

$\forall u \in S$

$$f(u) - f(\omega) \geq \frac{\sigma}{2} \|u - \omega\|^2$$



At the minima $f(u) - f(\omega)$ is growing by $\frac{\sigma}{2} \|u -$

This lemma holds for any convex function, need not be differentiable at all points. However can be easily derived for when function is differentiable at all points.

$$f'(\omega) = z = 0$$

$$\therefore f(u) \geq f(\omega) + \langle 0, u - \omega \rangle + \frac{\sigma}{2} \|u - \omega\|^2$$

$$\Rightarrow f(u) - f(\omega) \geq \frac{\sigma}{2} \|u - \omega\|^2$$

Test for convexity: 2nd derivative = +ve, for given function.

But when function is taking vectors as inputs, then Hessian should be positive semi-definite

---

<u>The test</u>:

$$\forall\, w, x \quad \langle \underbrace{\nabla^2 R(w)}_{\text{Hessian}}\, x, x \rangle \geq \sigma \|x\|^2$$

If true then function is σ - strongly convex

Checking now for <u>Euclidean Regularizer</u>

$$R(w) = \frac{1}{2}\|w\|_2^2$$

$$\nabla^2 R(w) = \begin{bmatrix} \frac{\partial^2 R(w)}{\partial w_1 \partial w_1} & \frac{\partial^2 R(w)}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 R(w)}{\partial w_1 \partial w_d} \\ \vdots & & & \\ \frac{\partial^2 R(w)}{\partial w_d \partial w_1} & \cdots & & \frac{\partial^2 R(w)}{\partial w_d \partial w_d} \end{bmatrix}$$

For the euclidean regularizer

$$= I$$

$$\therefore \langle \nabla^2 R(w)\, x, x \rangle$$

$$= \langle I x, x \rangle$$

$$= \langle x, x \rangle$$

$$= \|x\|_2^2$$

$\therefore$ $R(w)$ is 1 - strongly convex w.r.t. l_2 - norm

Entropy Regularizer

The update rule we derived with the L-2 regularizer, w

$$\omega_{t+1} = \omega_t - \eta z_t$$

But the ω_t's need not be probability vectors.

So in order to constraint to cases where ω_t has to be probability vectors, we need different regularizers.

KL-divergence

$$KL(p, q) = \sum_{i=1}^{k} p(i) \log \frac{p(i)}{q(i)}$$

$$= \sum p_i \log \frac{1}{q(i)} - \underbrace{\sum p(i) \log \frac{1}{p(i)}}_{\text{is called entropy}}$$

Thus entropy regularizer is:

$$\boxed{R(\omega) = \sum_{i=1}^{d} \omega_i \log \omega_i} \quad \omega \in S = \left\{ x \in \mathbb{R}^d \text{ s.t. } x_i \geq 0 \ \forall \ i = 1, 2 \ldots d, \ \sum_{i=1}^{d} x_i = 1 \right\}$$

Special case of with B = 1

S is thus a probability simplex / space

Taking the general case of S

$$S = \left\{ x \in \mathbb{R}^d : x_i \geq 0 \ \forall i, \ \|x_i\|_1 \leq B \right\}$$

In this case :

[R is $\frac{1}{B}$ - strongly convex w.r.t. l_1 - norm]

Check by finding out :

$$\langle \nabla^2 R(w) x , x \rangle \geq \frac{1}{B} \|x\|_1^2$$

FoReL with strongly convex functions

Lemma : Let $R : S \rightarrow \mathbb{R}$ be a σ - strongly convex function over S w.r.t. $\|.\|$

regularizer

Let : f_t is L-Lipschitz w.r.t $\|.\|$ $\forall t$.

If $w_1, w_2 \ldots$ are the predictions of FoReL then

$$\underbrace{f_t(w_t) - f_t(w_{t+1}) \leq L_t \| w_t - w_{t+1} \|}_{\text{already know this from the L-Lipschitz lecture}} \leq \frac{L_t^2}{\sigma}$$

this bound obtained for strongly convex regularizers