Lecture 15 - 04-05-2020

1.1 Regret analysis of OGD

We introduce the **Regret**.

$$\frac{1}{m}\sum_{t=1}^{T}\ell_t(w_t) - \frac{1}{T}\sum_{t=1}^{T}\ell_t(u_t^*)$$

$$(x_1, y_1)...(x_t, y_t) \qquad \ell_t(w) = \left(w^T x_t - y_t\right)^2$$

we build a loss function for example with the square loss.
The important thing is that $\ell_1, \ell_2, ...$ is a sequence of **convex losses**.

In general we define the regret in this way:

$$R_T(u) \;=\; \frac{1}{m}\sum_{t=1}^{T}\ell_t(w_t) - \frac{1}{T}\sum_{t=1}^{T}\ell_t(u_t)$$

The Gradiant descent is one of the simplest algorithm for minimising a convex function. We recall the iteration did by the algorithm:

$$w_{t+1} \leftarrow w_t - \eta_t \nabla f(w_t) \qquad \eta_t > 0 \ \textbf{learning rate} \quad f\ \textit{convex}$$

$$f : \mathbb{R}^d \rightarrow \mathbb{R}\ \textit{that's why use the gradiand instead of the derivative}$$

Learning rate can depend on time and we approach the region of the function f where the region is 0. We keep on moving in the X axes in the direction where the function is decreasing.

1.1.1 Projected OGD

2 parameters: $\eta > 0$ and $U > 0$
Initialisation: $w_1 = (0, ..., 0)$
For $t = 1, 2, ...$

1) **Gradiant step**:

$$w'_{t+1} = w_t - \frac{\eta}{\sqrt{t}} \nabla \ell_t(w_t) \qquad (x_t, y_t) \rightsquigarrow \ell_t$$

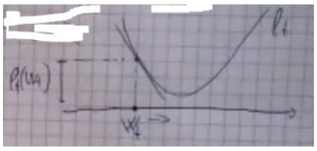

Figure 1.1:

2) **Projection step**:

$$w_{t+1} = arg \min_{w:\|w\| \le U} \|w - w'_{t+1}\|$$

Projection of w'_{t+1} onto the ball of radius U.

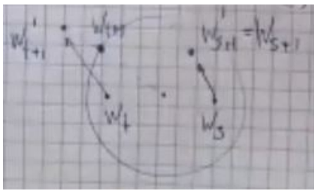

Figure 1.2:

Now we define the Regret:

$$U^*_T = arg \min_{U \in \mathbb{R}^d \ \|U\| \le U} \frac{1}{T} \sum_{t=1}^{T} \ell_t(U)$$

We are interested in bounding the regret $R_T\left(U^*_T\right)$

I will Fix $\ell_1, ... \ell_t$ let $U = U^*_T$ for U.
Taylor's theorem for multivariate functions

Let's look a univariate first $\quad f : \mathbb{R} \to \mathbb{R}$ *(has to be twice differentiable)*
$w, u \in \mathbb{R}$

$$f(u) = f(w) + f'(w)\,(u - w) + \frac{1}{2}\,f''(x)\,(u - w)^2$$

For the multivariate case:
$f : \mathbb{R}^d \to \mathbb{R}$ twice differentiable $\forall u, w \in \mathbb{R}^d$

$$f(u) = f(w) + \nabla f(w)^T\,(u - w) + \frac{1}{2}\,(u - w)^T\,\nabla^2 f(\xi)\,(u - w)$$

where ξ is some point on the segment goining u and w. We have the Hessian matrix of f:

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i\,\partial x_j}|x = x_i$$

If f is convex then, $\nabla^2 f$ is positive and semidefinite.
$\forall x \in \mathbb{R}^d \quad \forall z \in \mathbb{R}^d \qquad z^T\,\nabla^2 f(x)\,z \geq 0$

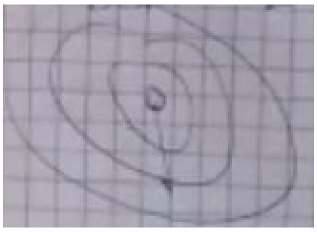

Figure 1.3:

min. 38:38