CMPUT 367

Roderick Lan

Contents

Lecture 6 - Jan 30

1 Validation

Train set $\leftarrow$ minimize objective
Validation set $\leftarrow$ hyperparameter selection (try model internally)
Test set $\leftarrow$ mimic deployment of ML system

For diff. values of λ_1:

> Train $h^* = \arg\min_{h \in \mathcal{H}} J_\lambda(h, \mathcal{D}_{\text{train}})$
>
> Validate $\text{err}_\lambda = \text{err}(\lambda) = \text{Err}(h^*_\lambda, \mathcal{D}_{\text{val}})$
> error function is "client's" error

Choose $\lambda^* = \arg\min_\lambda \text{err}(\lambda) \leftarrow$ delivery of sys. to client
(maximize over previous λs)

Report $\text{Err}(h^*_{\lambda^*}, \mathcal{D}_{\text{test}}) \leftarrow$ client's disatisfaction

What if we have multiple hyperparams (ie. $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$) each taking v values.
size: $|v|^k$
Minimize with coordinate descent (fix all but one λ, tune λ; repeat for all).
Can't do regular gradient descent; ($\arg\min f$) nondifferentiable

1.1 Why not learn hyperparam during training?

$$\arg\min_{w,\lambda} \frac{1}{2M}\|Xw - t\|^2 + \lambda\|w\|^2 \Rightarrow \lambda^* = 0$$

Ineffective

2 Convexity

(defn in lec 5)

Definition 2.0.1: Concave

f is a concave function iff $-f$ is a convex function
(not important)

Convex functions may not be differentiable; but must be continuous
Develop conditions similar to convexity that are easy to verify. (convexity hard to verify for all points)

Thm 2.0.1

Let f be a twice-differentiable function on a convex domain.
The following three statements are equivalent:

1. f is a convex function
2. [1st order condition]
   $\forall x, y \in \text{domain}_f$
   $$f(y) \geq f(x) + [\nabla_x f(x)]^\top (y - x)$$
   $[\nabla_x f(x)]^\top (y - x) = \sum_{i=1}^{d} \frac{\partial f(x)}{\partial x_i}(y_i - x_i)$
   (tangent line/hyperplane)
3. [2nd order condition]
   $\forall x \in \text{dom}_f$
   $$\begin{aligned} H &= \nabla_x^2 f(x) \geq 0 \\ &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_d} \end{bmatrix} \end{aligned}$$

$$\begin{aligned} f(x_1, x_2) &= x_1^2 + x_2^2 \\ \frac{\partial f}{\partial x} &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{bmatrix} \end{aligned}$$

zero diagonal + pos and neg diagonal $\rightarrow$ disentangled
nonzero offdiagonal $\rightarrow$ entangled; need to find eigenvecs/eigenvals
If H is not diagonal, then you need to check eigenvalues
All eigenvalues nonnegative $\iff$ Positive Semi-Definite (PSD)