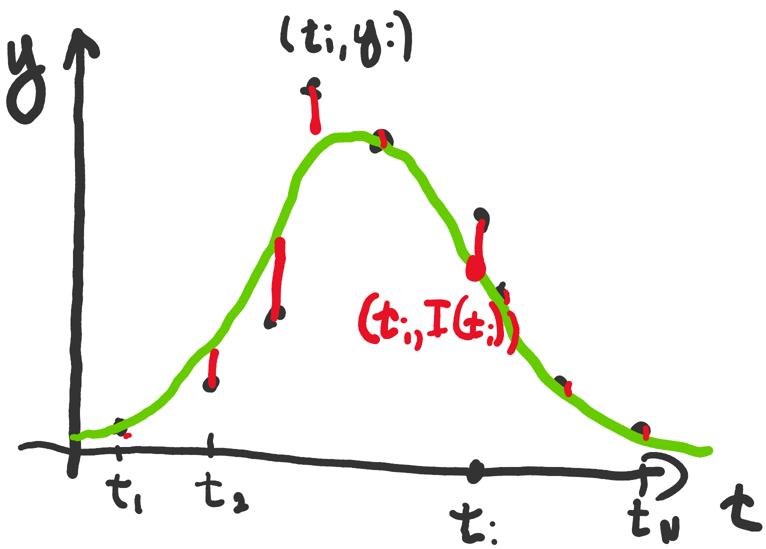
Monday, August 3, 2020 12:53 PM

TODAY: • FITTING ODES WITH GRAD. DECENT

• INTRO TO PDES

LOSS MINIMIZATION FRAMEWORK:

GIVEN DATA POINTS (t_i, y_i), $i = 1, \ldots, N$



PROPOSE A FUNCTIONAL MODEL:

$$S' = -\beta S I$$

$$I' = \beta S I - \gamma I$$

GIVES FUNCTIONS $S(\underset{\text{VAR}}{t} \mid \underset{\text{PARAMS}}{\beta, \gamma})$ AND $I(t \mid \beta, \gamma)$.

FOR SOME β, γ, WANT TO EVALUATE HOW CLOSE OUR MODEL IS TO DATA.

DEFINE AN ERROR TO BE SOME

DEFINE AN ERROR TO BE SOME MEASURE OF DIFFERENCE BETWEEN DATA AND MODEL.

COMMON <u>ERRORS</u> OR <u>LOSS FUNCTIONS</u>

MEAN SQUARED ERROR:

$$MSE(\beta,\gamma) = \frac{1}{N}\sum_{i=1}^{N}\Big(\underbrace{y_i}_{\text{ACTUAL VALUE}} - \underbrace{I(t_i)}_{\text{PREDICTED VALUE}}\Big)^2$$

MEAN ABSOLUTE ERROR:

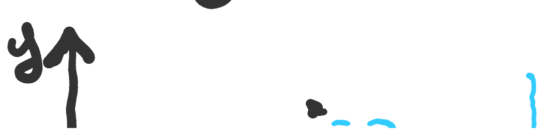
$$MAE(\beta,\gamma) = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - I(t_i)\right|$$

GOAL: FIND (β,γ) THAT MINIMIZE OUR CHOSEN ERROR.

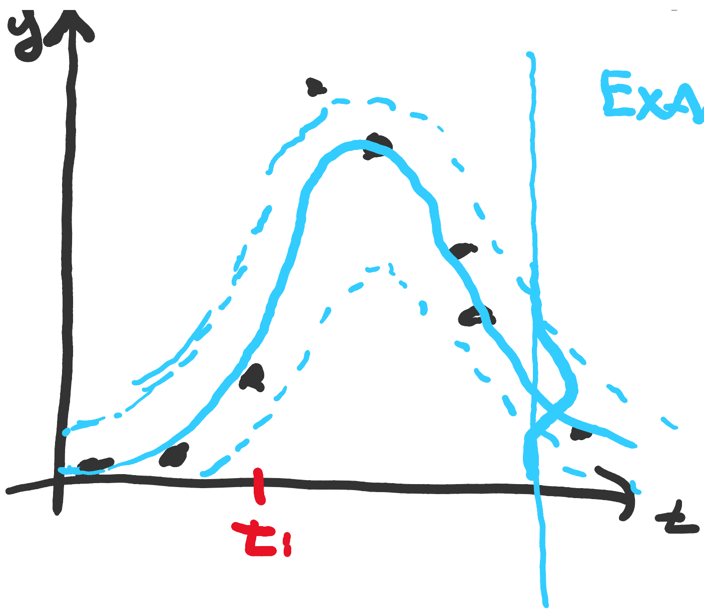
MAXIMUM LIKELIHOOD FRAMEWORK

AGAIN WE HAVE DATA (t_i, y_i), $i=1,\ldots,N$.

PROPOSE A <u>PROBABALISTIC</u> MODEL: GIVEN ANY t_i, WE HAVE A PROB. DISTRIBUTION ON y_i AT t_i: $P(y|t)$ ("PROB. OF y GIVEN t")



EXAMPLE: GAUSSIAN



EXAMPLE: GAUSSIAN DIST. AROUND A FUNCTION:

$$P(y|t) = \mathcal{N}(y | I(t), \sigma)$$

$I(t)$ = μ · MEAN

GOAL: MAXIMIZE THE LIKELIHOOD DATA WAS <u>ACTUALLY</u> DRAWN FROM P.

LIKELIHOOD (t_i, y_i) CAME FROM P IS JUST THE PDF:

$$L_i = P(y_i | t_i)$$

IN EX:

$$P(y_i | t_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(y_i - I(t_i))^2}{2\sigma^2} \right)$$

<u>TOTAL LIKELIHOOD:</u> PRODUCT OF LIKELIHOODS

$$L = \prod_{i=1}^{N} L_i$$, ASSUMING NOISE IS INDEPENDENT

<u>EX:</u>

$$L(\ldots) = \prod^{N} \ldots \exp\left(\frac{(y_i - I(t_i))^2}{\ldots} \right)$$

EX,

$$L(\beta,\gamma,\sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - I(t_i))^2}{2\sigma^2}\right)$$

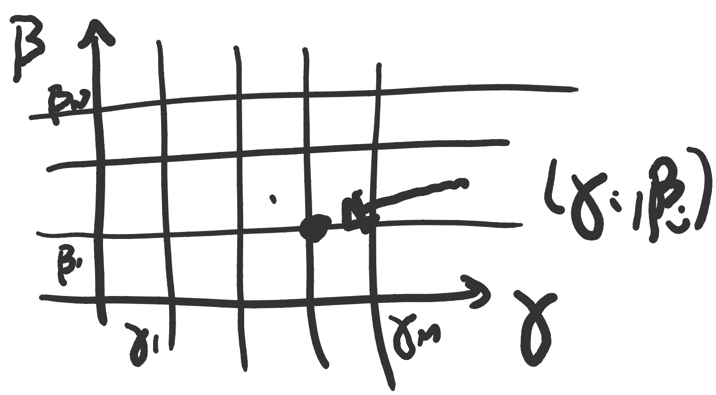
NOTICE: WANT TO MAXIMIZE, FIX σ BY FINDING S.T.D. OF $\{y_i - I(t_i)\}$, THEN

$$L(\beta,\gamma,\sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\sum_{i=1}^{N} \frac{(y_i - I(t_i))^2}{2\sigma^2}\right)$$

SO MAXIMIZING LIKELIHOOD FOR A MODEL W/ GAUSSIAN NOISE IS SAME AS MINIMIZING <u>MSE</u>.

HOW DO WE MINIMIZE LOSS?

- GRID SEARCH:



CHECK MSE FOR EACH (γ_i, β_j), FIND MIN.

MIN.

PROBLEMS:

- GRIDE SIZE MAY MISS POINTS
- HARD TO SEE REPEATING BEHAVIOR
- COMPUTATIONALLY EXPENSIVE: FOR P PARAMS WITH V VALUES:

  $$O(V^P)$$

- AND MOST VALUES ARE NO WHERE NEAR INTERESTING REGIONS.

BUT!

LOTS OF IMPROVEMENTS

- ZOOMING
- Log
- STOCHASTIC ← VERY USEFUL
- WORKS FOR NON DIFFERENTIABLE FUNCTIONS

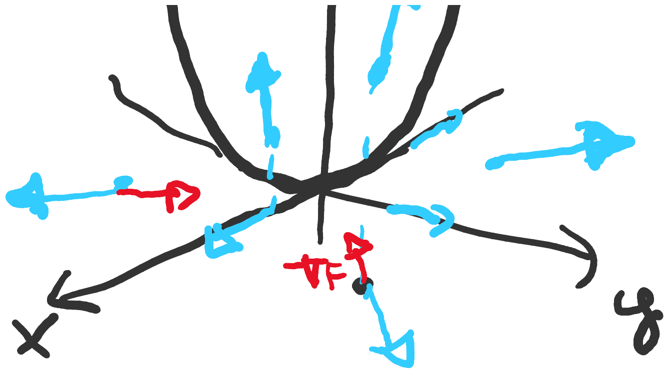
- GRADIENT DECENT:

ASSUME WANT TO MINIMIZE $F(x,y)=x^2+y^2$

$z=F(x,y)$

RECALL:

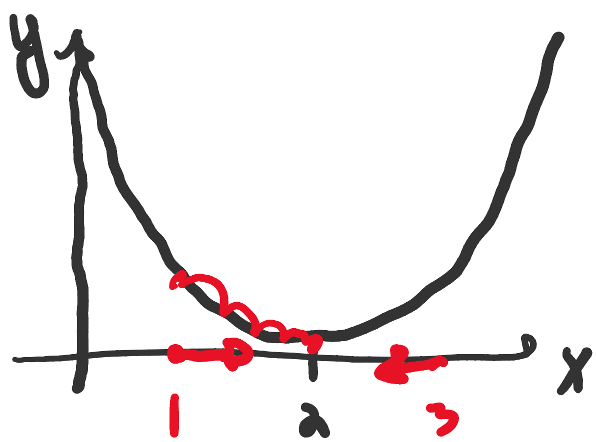
$\nabla F(x,y)$ POINTS



$\nabla F(x,y)$ POINTS IN DIR OF GREATEST INCREASE OF F.

Ex: $\nabla F = \langle 2x, 2y \rangle$. So $\nabla F(1,1) = \langle 2,2 \rangle$

THIS MEANS IF WE WANT TO FIND A MINIMUM, WALK OPP. DIRECTION $-\nabla F$.



$y = (x-2)^2 = F$

$\nabla F = 2(x-2)$

$-\nabla F = 2(x-2)$

FOLLOWING $-\nabla F$, BOUNCE DOWN TOWARDS LOCAL MINIMUM.

☐ GRADIENT DECENT ALGORITHM:

TO FIND LOCAL MINIMUM FOR $F(\vec{x})$, FIX STARTING LOCATION $\vec{x}_0$, FIX STEP SIZE OR LEARNING RATE

FIX STARTING LOCATION x_0, AND STEP SIZE OR LEARNING RATE λ.

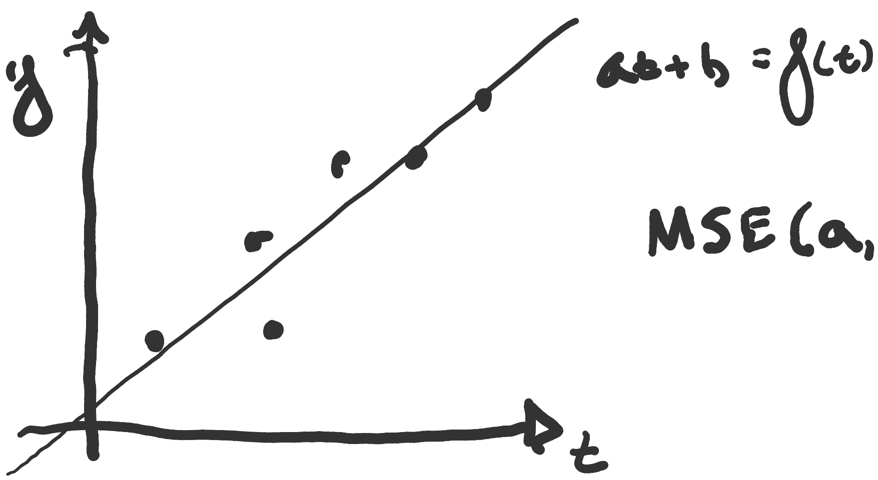
THEN:

$$\vec{x}_n = \vec{x}_{n-1} - \lambda \nabla F(\vec{x}_{n-1})$$

STOP WHEN $\|\nabla F\| < \varepsilon$, FOR SOME FIXED $\varepsilon > 0$.

APPLY TO LOSS FUNCTIONS:

<u>Ex</u>: LINEAR MODEL WITH MSE



$$MSE(a,b) = \frac{1}{N}\sum_{i=1}^{N}\Big(y_i - \underbrace{at_i - b}_{\hat{y}(t_i)}\Big)^2$$

WANT TO OPTIMIZE MSE WITH RESPECT TO a AND b: NEED TO COMPUTE

$$\nabla MSE = \left\langle \frac{\partial MSE}{\partial a}, \frac{\partial MSE}{\partial b} \right\rangle$$

$$= \left\langle \frac{2}{N}\sum_{i=1}^{N} -t_i(y_i - at_i - b), -\frac{2}{N}\sum_{i=1}^{N}(y_i - at_i - b) \right\rangle$$

$$= \left\langle \frac{2}{N}\sum_{i=1} -t_i(y_i - at_i - b), \frac{2}{N}\sum_{i=1} (y_i - a_i t_i - b) \right\rangle$$

DOES NOT DEPEND ON t OR y, ONLY a AND b.

THEN

$$\langle a_n, b_n \rangle = \langle a_{n-1}, b_{n-1} \rangle - \lambda \nabla MSE(a_{n-1}, b_{n-1})$$

GRADIENT DECENT FOR ODES

IN ALL EXAMPLES ABOVE WE HAVE AN EXPLICITLY DEFINED FUNCTIONAL MODEL, LIKE $g(x) = x^2$. FOR MOST ODES WE DON'T. FOR EXAMPLE

$$S' = \beta SI$$
$$I' = \beta SI - \gamma I$$

NO CLOSED FORM EXPRESSION FOR S, I. SO WE MUST ESTIMATE GRADIENT USING ODE INT:

$$\frac{\partial L}{\partial \beta} \approx \frac{L(\beta + \Delta\beta, \gamma) - L(\beta, \gamma)}{\Delta \beta}$$ NUMERICAL ESTIMATE.

$$\frac{\partial}{\partial \beta} \approx \frac{\cdots}{\Delta \beta}$$ ESTIMATE.

CAN ESTIMATE ∇L, NOTE ODES SHOULD BE SMOOTH SO SHOULD BE REASONABLE FOR $\Delta\beta \ll 1$.

<u>PLAN</u>: GIVEN DATA (t_i, y_i) AND A MODEL,

$$S' = -\beta S I$$
$$I' = \beta S I - \gamma I$$

PICK STARTING PARAMS, $\beta_0, \gamma_0, I_0, S_0$,

GIVEN $(\beta_n, \gamma_n, I_n, S_n)$,

- NUMERIC SOLUTION USING ODEINT
- USE NUMERICAL SOLUTION TO COMPUTE LOSS

PYTHON/SCIPY

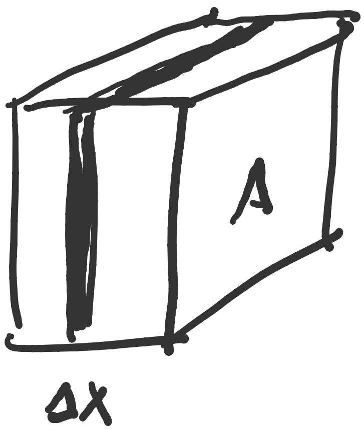
- REPEAT FOR $\beta_n + \Delta\beta$, $\gamma_n + \Delta\gamma$, ETC
- ESTIMATE ∇L

SKIP

- ESTIMATE ∇L
- USE GRAD DECENT: $(\beta_{n+1}, \dots) = (\beta_n, \dots) - \nabla L$

PARTIAL DIFFERENTIAL EQUATIONS:

RECALL HEAT THROUGH A WALL:



$$\left\{ \text{RATE OF HEAT CHANGE IN SECTION} \right\} = \left\{ \text{RATE HEAT IN} \right\} - \left\{ \text{RATE HEAT OUT} \right\}$$

LETS MOVE AWAY FROM THERMAL EQ.

LET $Q = cm \frac{\partial u}{\partial t}$, THEN HEAT FLOW IS

$$cm \frac{\partial u}{\partial t} = J(x,t)A - J(x+\Delta x, t)$$

LET ρ BE WALL DENSITY, SO $m = \rho A \Delta x$, THEN

THEN

$$c\rho A\Delta x \frac{\partial u}{\partial t} = J(x,t)A - J(x+\Delta x, t)A$$

OR

$$c\rho \frac{\partial u}{\partial t} = -\frac{J(x+\Delta x, t) - J(x,t)}{\Delta x}$$

IN LIMIT:

$$c\rho \frac{\partial u}{\partial t} = -\frac{\partial J}{\partial x} = K\frac{\partial^2 u}{\partial x^2}$$

SIMPLIFY TO

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}$$ HEAT EQUATION

THIS IS A PARTIAL DIFF. EQ.

SOLVING HEAT EQUATION USING SEPERATION OF VARIABLES

ASSUME WE CAN WRITE

ASSUME WE CAN WRITE

$$U(x,t) = \phi(x)\, G(t)$$

FOR $\phi, G : \mathbb{R} \to \mathbb{R}$ ONE VARIABLE.

THEN:

$$\frac{\partial U}{\partial t} = \phi(x)\, G'(t) = \alpha\, \phi''(x)\, G(t) = \frac{\partial^2 U}{\partial x^2}$$

AND SO

$$\frac{1}{G(t)} \frac{\partial G}{\partial t} = \alpha \frac{1}{\phi(x)} \frac{\partial^2 \phi}{\partial x^2}.$$

HOW CAN A FUNCTION OF x EQUAL A FUNCTION OF t? BOTH <u>MUST</u> BE CONSTANT. THEN CAN SOLVE

(1) $\alpha\lambda = \frac{1}{G} \frac{\partial G}{\partial t}$

(2) $\lambda = \frac{1}{\phi} \frac{\partial^2 \phi}{\partial x^2}$

FOR (1),

FOR (1),

$$\int \alpha\lambda \, dt = \int \frac{1}{G} \, dG \Rightarrow G = Ae^{\alpha\lambda t}$$

FOR (2), REWRITE

$$\phi'' - \lambda \phi = 0$$

CHAR. POLY: $r^2 - \lambda = 0$

SO $r = \pm\sqrt{\lambda}$

AND

$$\phi = c_1 e^{\sqrt{\lambda} x} + c_2 e^{-\sqrt{\lambda} x}$$

SO

$$u(t,x) = \left(c_1 e^{\sqrt{\lambda} x} + c_2 e^{-\sqrt{\lambda} x}\right) e^{\alpha\lambda t}$$

FOR PDE'S BOUNDARY CONDITIONS DETERMINE <u>EVERYTHING</u>.

DETERMINE EVERYTHING.

EX: HEAT EQUATION W/ HOMOGENEOUS BOUNDARY CONDITIONS

ASSUME $U(0,t) = U(L,t) = 0$. THEN

$$\phi(0) = 0 = \phi(L)$$

AND SO

$$0 = C_1 + C_2 \Rightarrow C_1 = -C_2$$

AND

$$0 = C_1 e^{\sqrt{\lambda} L} - C_1 e^{-\sqrt{\lambda} L} = C_1 \left(1 - e^{-2\sqrt{\lambda} L}\right)$$

SO

$$1 = e^{-2\sqrt{\lambda} L}$$

WHICH IS NOT POSSIBLE FOR $L, \sqrt{\lambda} > 0$. BUT WHAT IF $\lambda < 0$? THEN

$$1 = \cos 2\sqrt{|\lambda|} L - i \sin 2\sqrt{|\lambda|} L$$

HOLDS EXACTLY WHEN $|\lambda| = \left(\frac{\pi n}{L}\right)^2$

FOR $n \in \mathbb{Z}$. SO HAVE INFITE BUT DISCRETE # OF SOLUTIONS:

$$U(t,x) = A\cos\left(\frac{2\pi n}{L}x\right)e^{-\alpha\left(\frac{\pi n}{L}\right)^2 t}$$