CS 180 Discussion 10

Diffusion and Flow matching

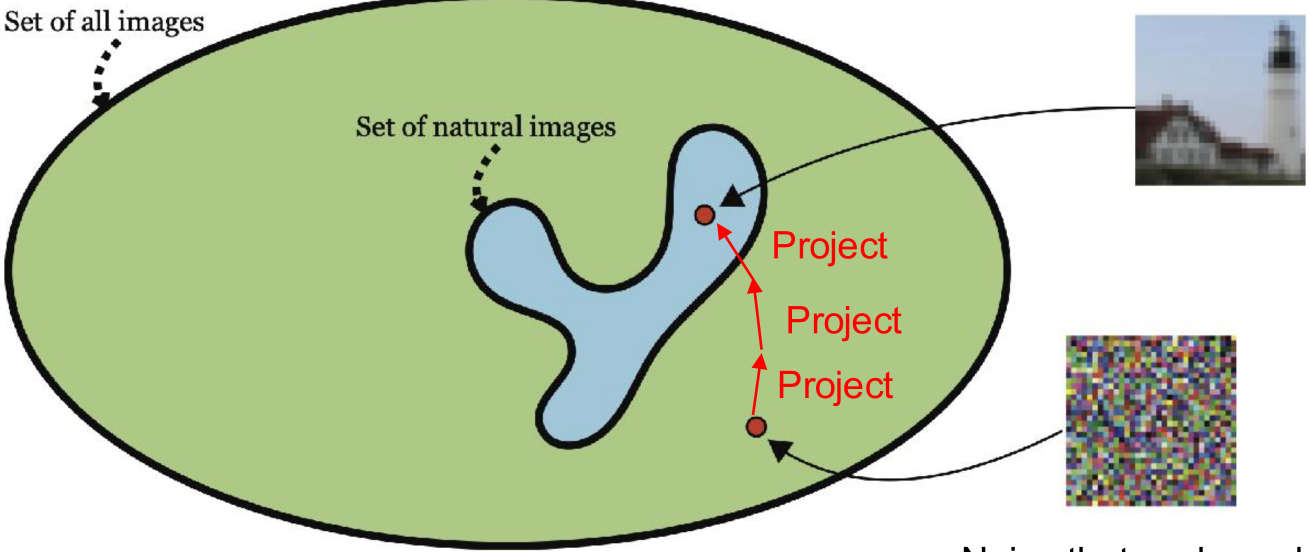
Agenda

1. How diffusion / flow matching works?
2. How CFG works?

Generating images by “projecting” onto the manifold

“Projection” to the natural image manifold

Nice image that we wish to have
But cannot generate



Noise that we know how to generate

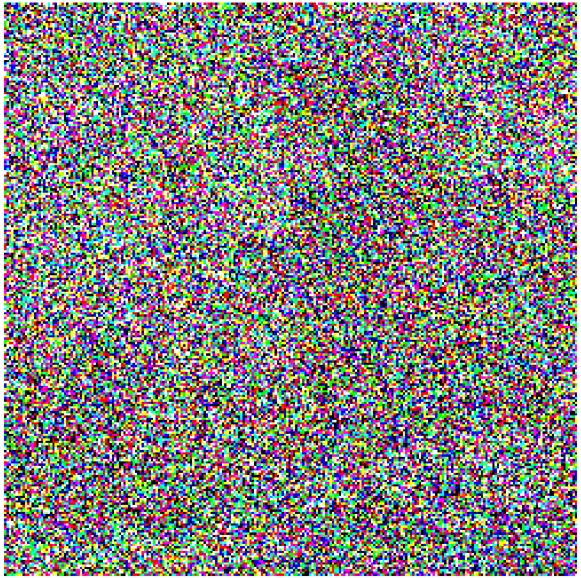
Diffusion process makes the “projecting” possible!

(Make it real)

Diffusion: “Project” = “denoise”

Forward (adding noise)

Reverse (denoising / generating)



Song 2021

How to properly add noise and denoise?

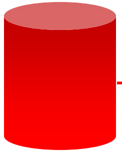
(If it’s that easy we wouldn’t wait until 2020 to make it work)

*Diffusion model was first discovered by Sohl Dickstein in 2015

How to add noise?

How to add noise?

Dataset

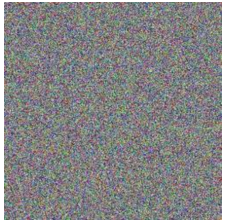


Sample

(H, W, 3)

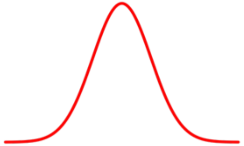
x_{clean}

(H, W, 3)



x_{noise}

Sample



$\mathcal{N}(0, I)$

High dimension

$t \sim [0, 1]$ Sample 0.9

Interpolation! $x_t = t x_{\text{clean}} + (1 - t) x_{\text{noise}}$ "Flow" style

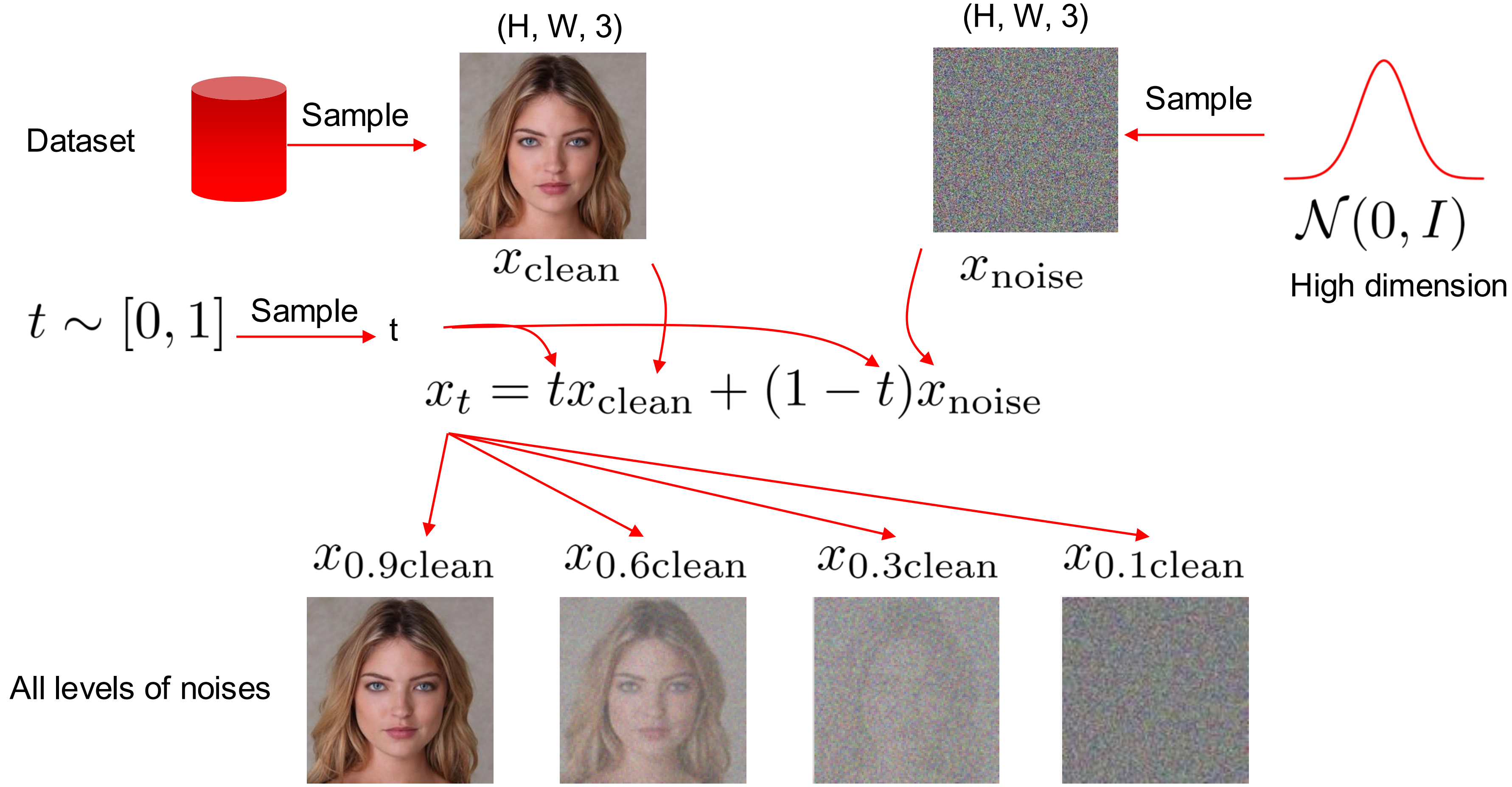
$x_t = \sqrt{\bar{\alpha}_t} x_{\text{clean}} + \sqrt{1 - \bar{\alpha}_t} x_{\text{noise}}$

"Diffusion" style
The confusing part:
t = 0 = clean!
t = a lot = noise!
(be cautious!)

$x_{0.9\text{clean}}$

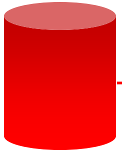
*We will use "Flow" style here!

How to add noise?



How to add noise?

Dataset

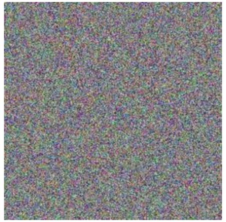


Sample

(H, W, 3)

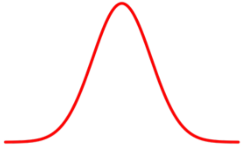
x_{clean}

(H, W, 3)



x_{noise}

Sample



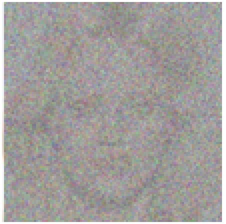
$\mathcal{N}(0, I)$

High dimension

$t \sim [0, 1] \xrightarrow{\text{Sample}} 0.3$

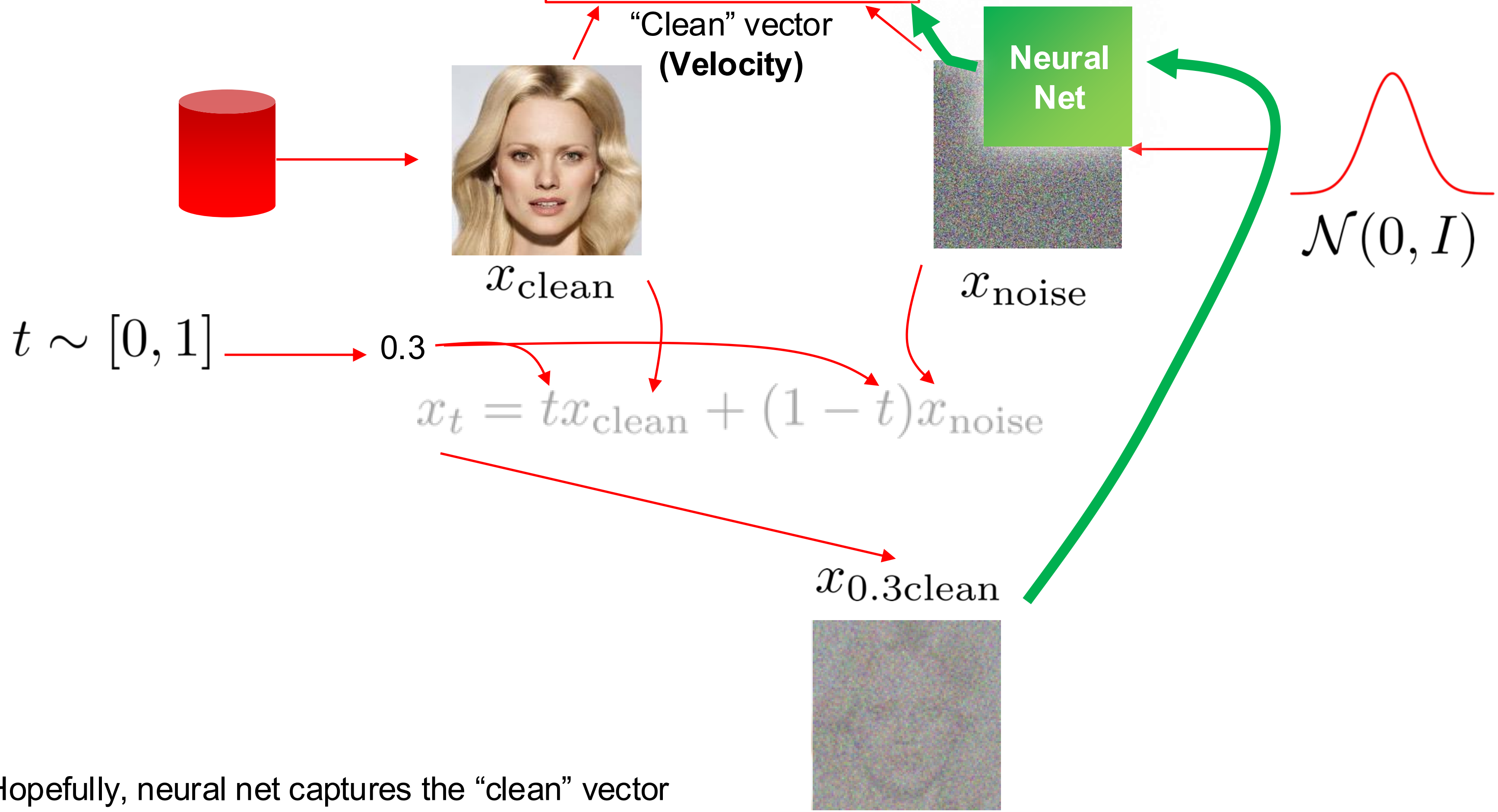
$$x_t = t x_{\text{clean}} + (1 - t) x_{\text{noise}}$$

$x_{0.3\text{clean}}$



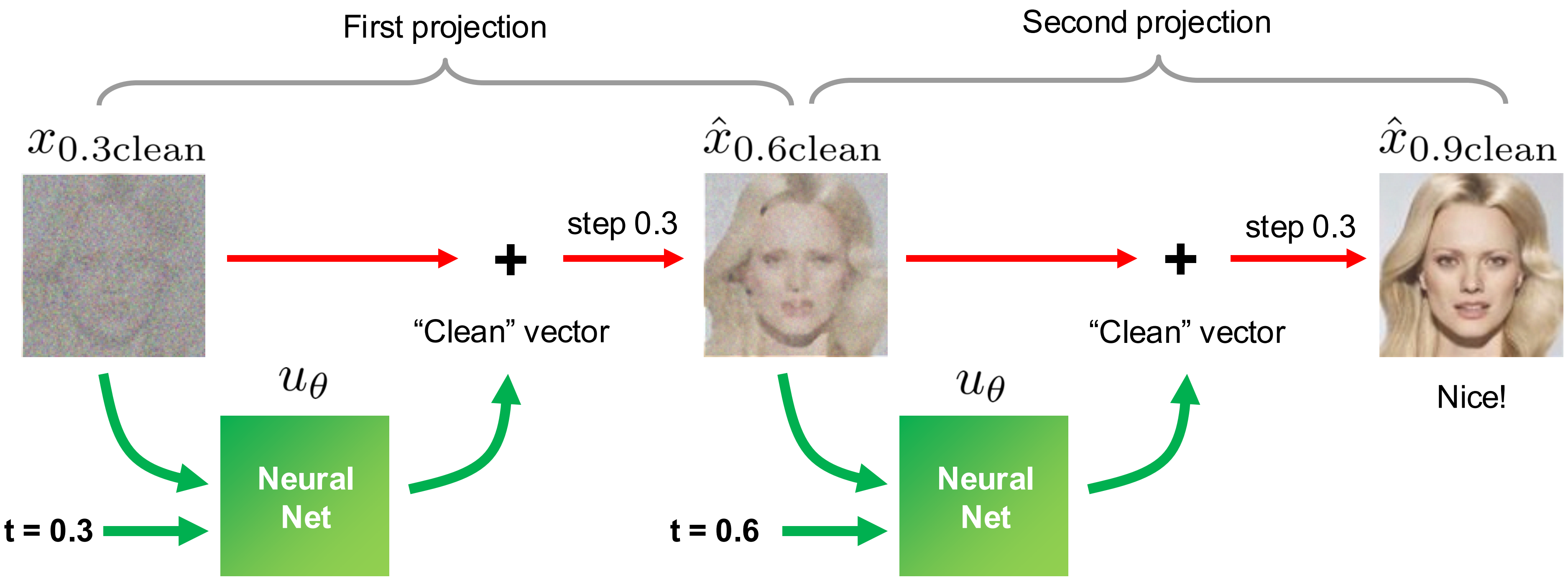
Learn to denoise…

Flow matching loss: $L = ||\boxed{(x_{\text{clean}} - x_{\text{noise}})} - u_\theta(x_t)||^2$



*Hopefully, neural net captures the "clean" vector

Now, how to project? Ideas?



*Subtle but useful

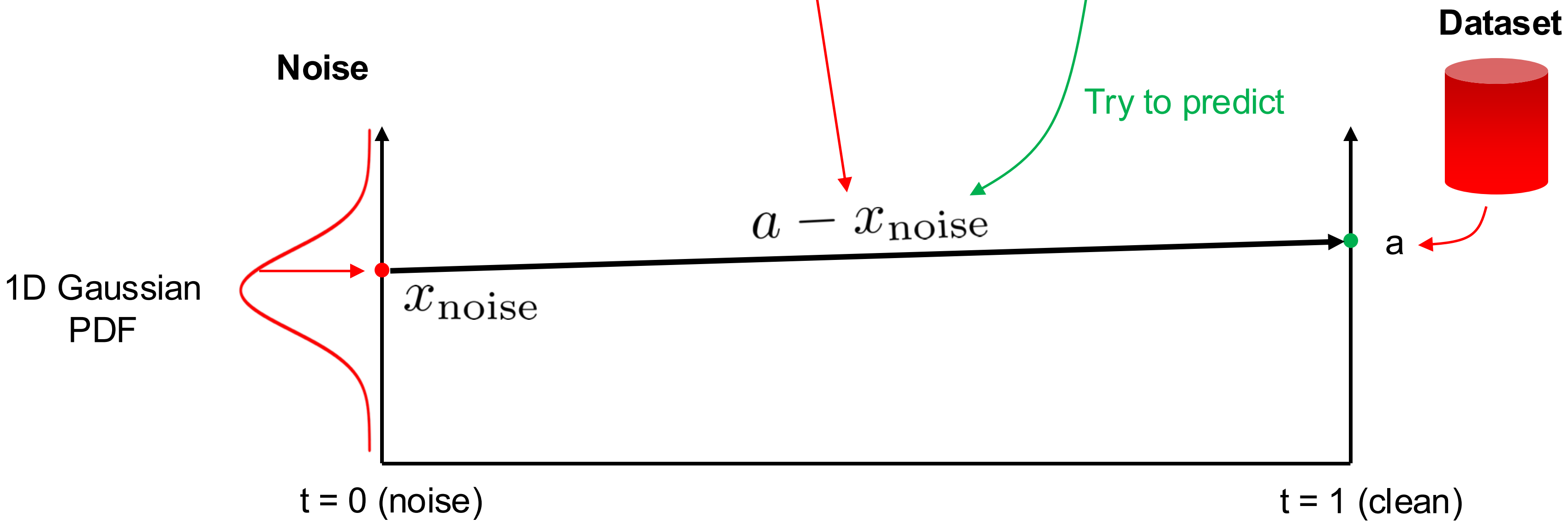
*Adjustable step size

To really understand what it learns, Imagine a 1D world

1D world with 2 datapoints

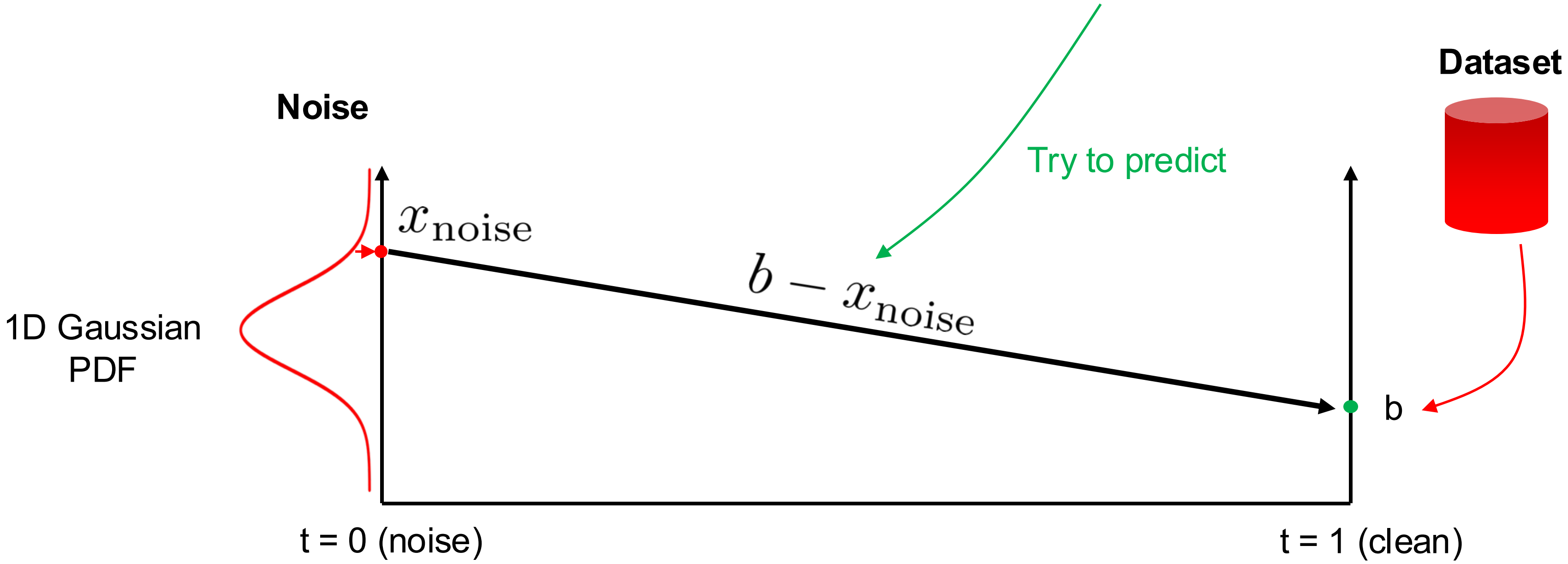
Sample noise, data, optimize

$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_{\text{noise}})||^2$$



Sample noise, data, optimize

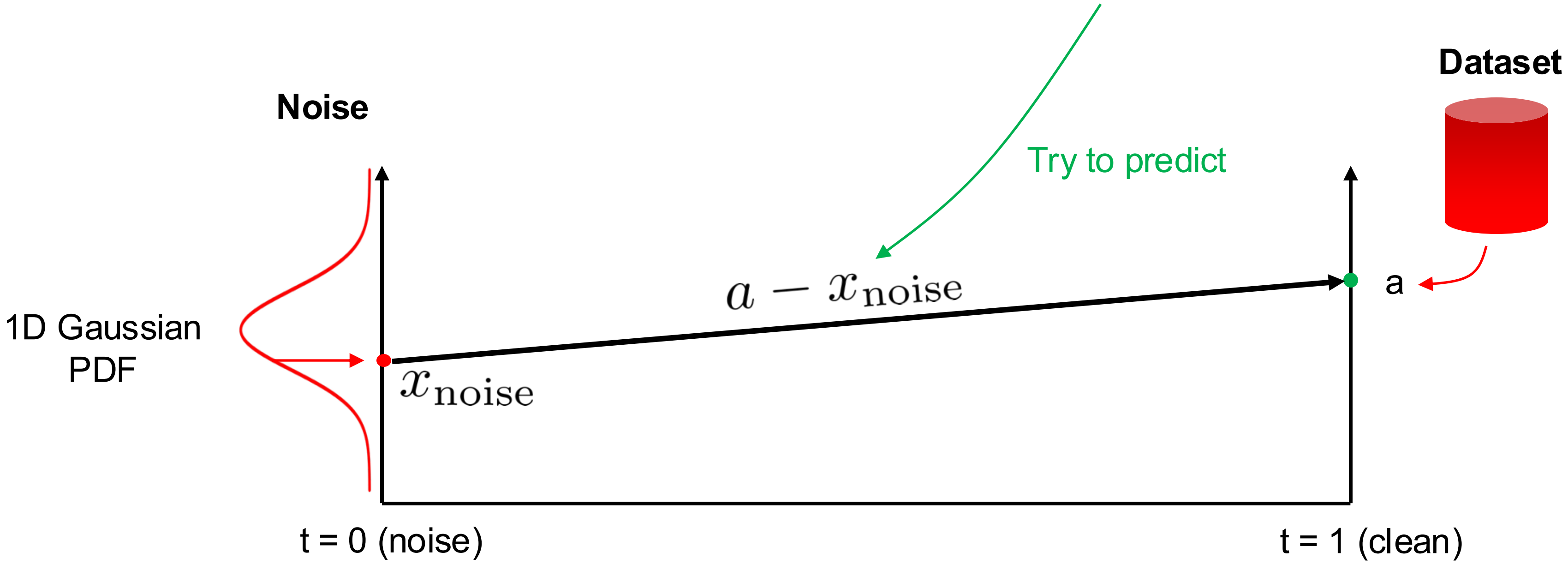
$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_{\text{noise}})||^2$$



*Sample many times…

Sample noise, data, optimize

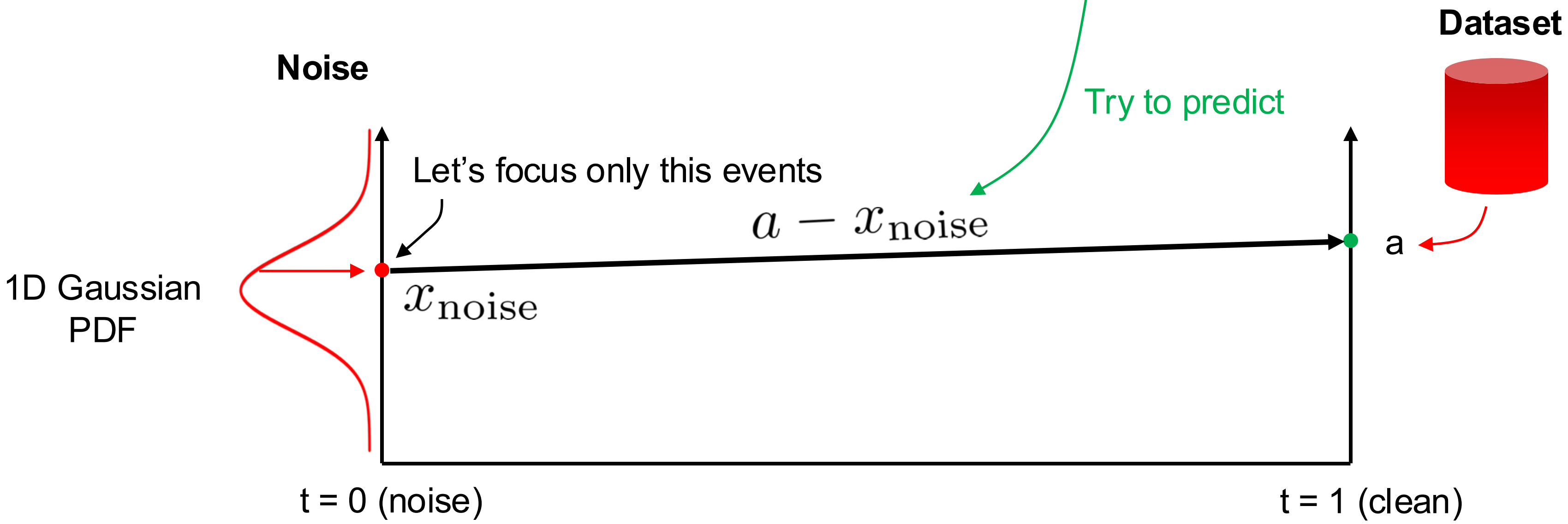
$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_{\text{noise}})||^2$$



*Sample many times…

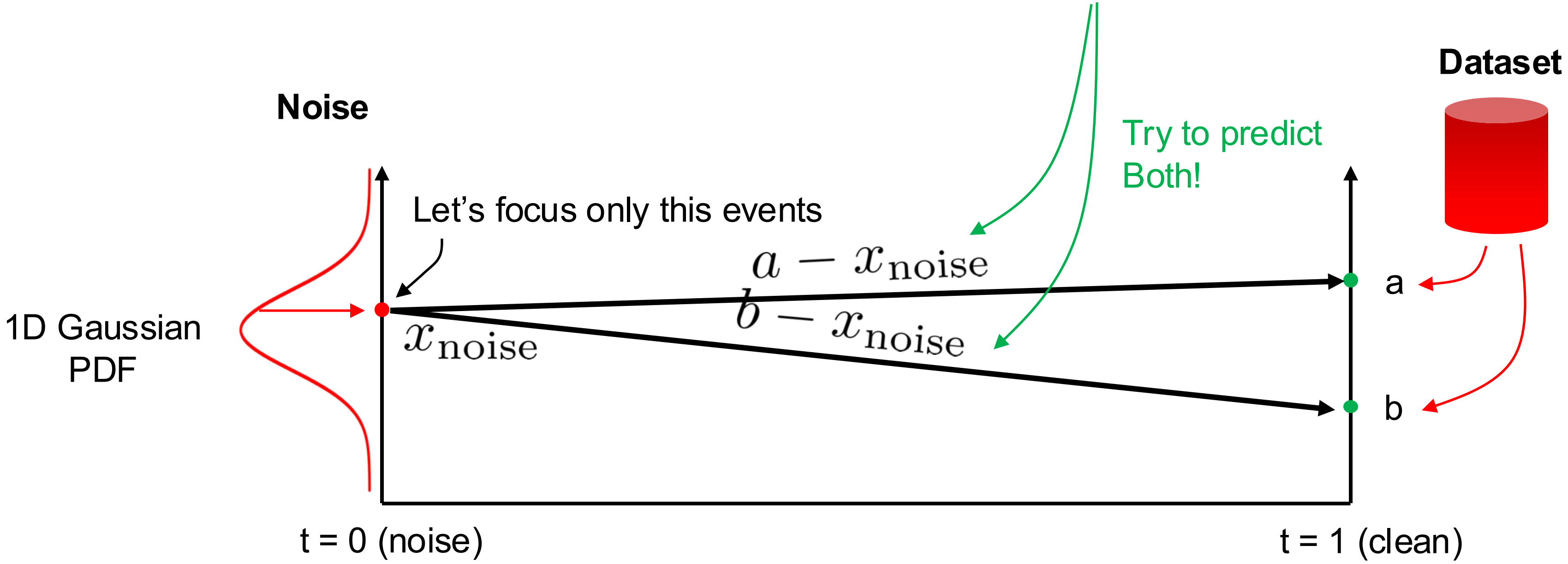
Let's focus on one event!

$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_{\text{noise}})||^2$$



We have two datapoints!

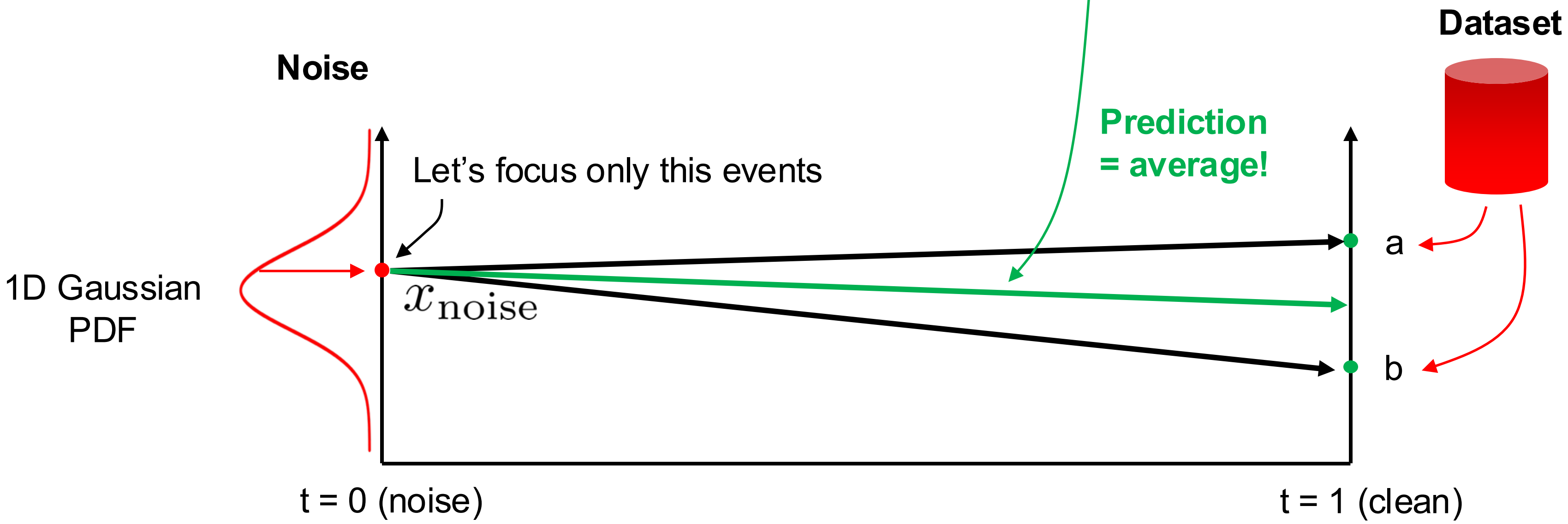
$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_{\text{noise}})||^2$$



*Can neural net predict two values?

What will neural net predict? (Marginal flow)

$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_{\text{noise}})||^2$$

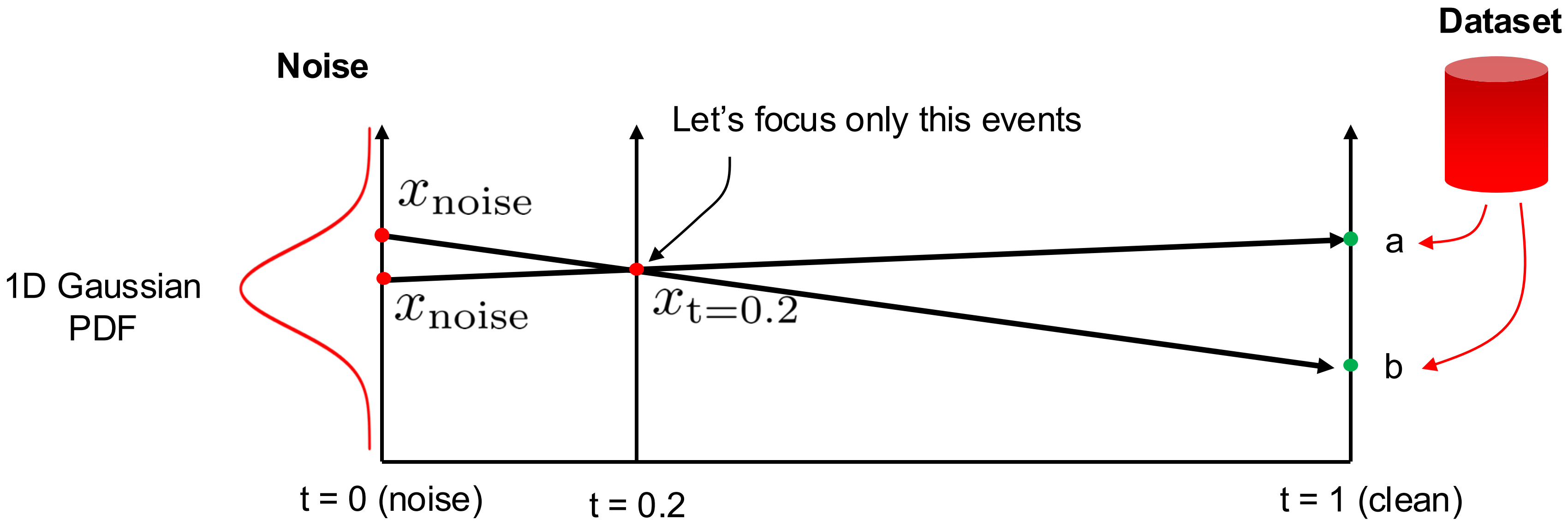


*At noise, the prediction is always the average!

Let’s try at $t > 0$

What will neural net predict? (Marginal flow)

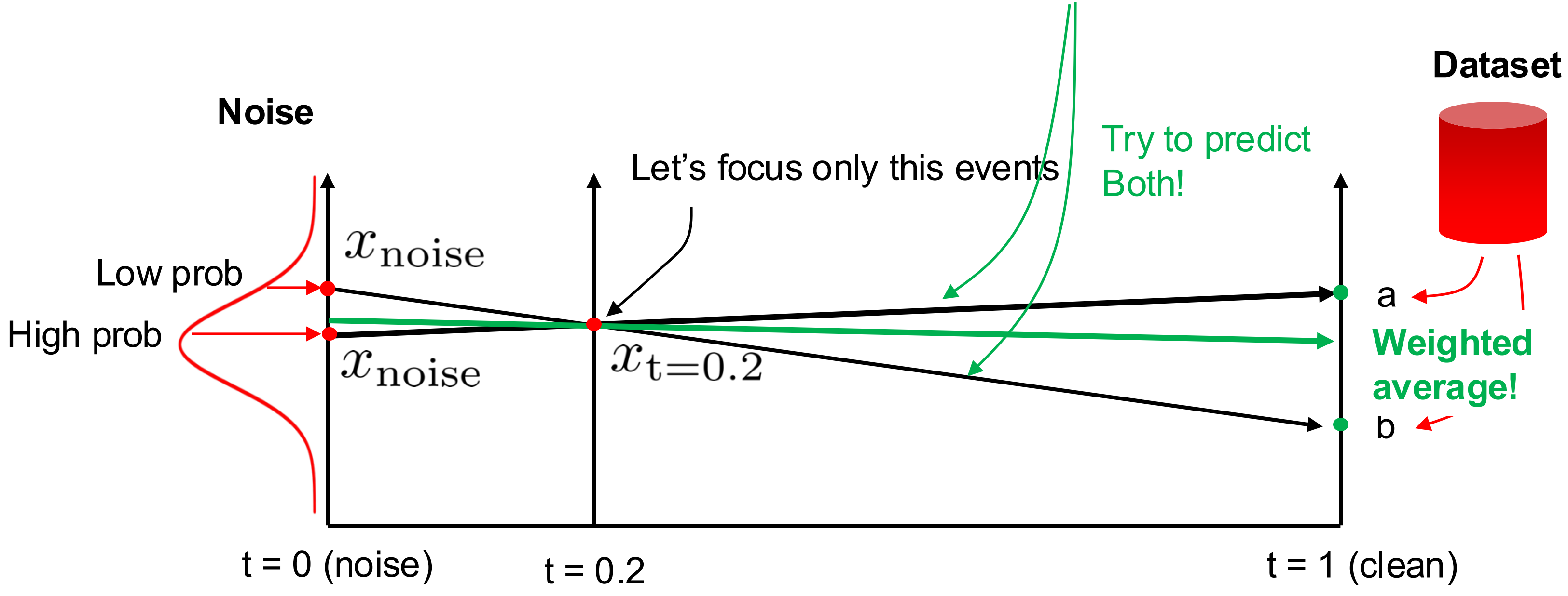
$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_t)||^2$$



Don't forget: $x_t = tx_{\text{clean}} + (1-t)x_{\text{noise}}$

What will neural net predict? (Marginal flow)

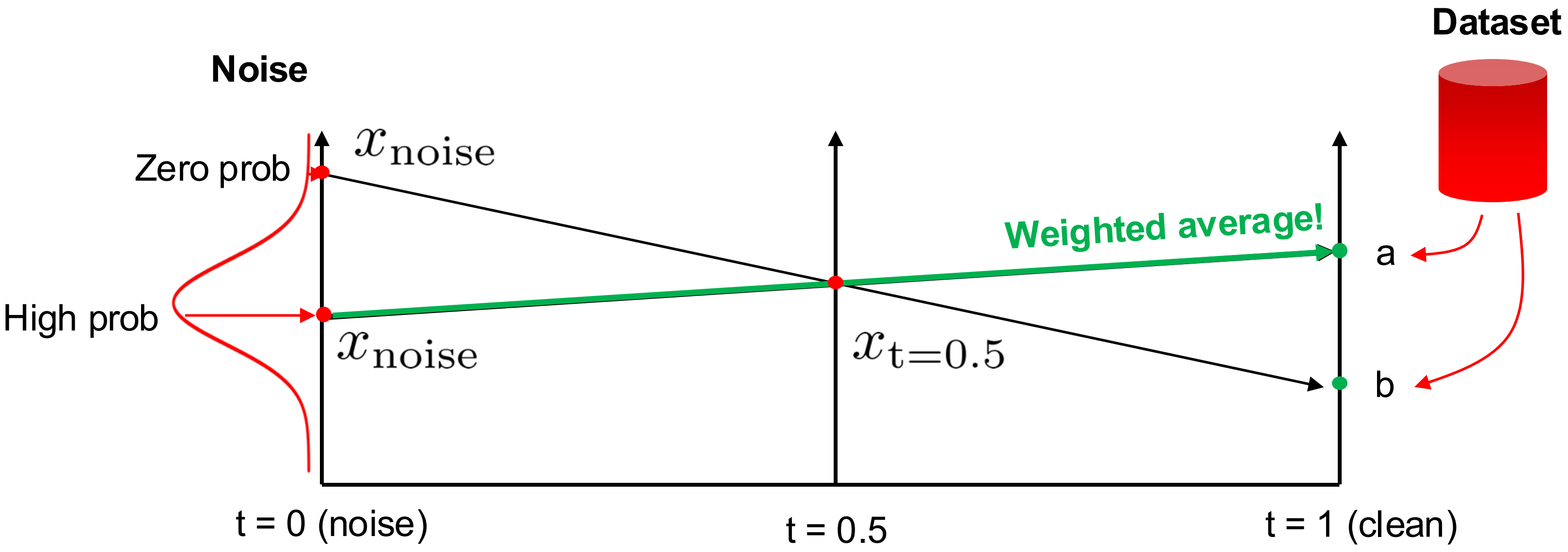
$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_t)||^2$$



Don't forget: $x_t = t x_{\text{clean}} + (1-t) x_{\text{noise}}$

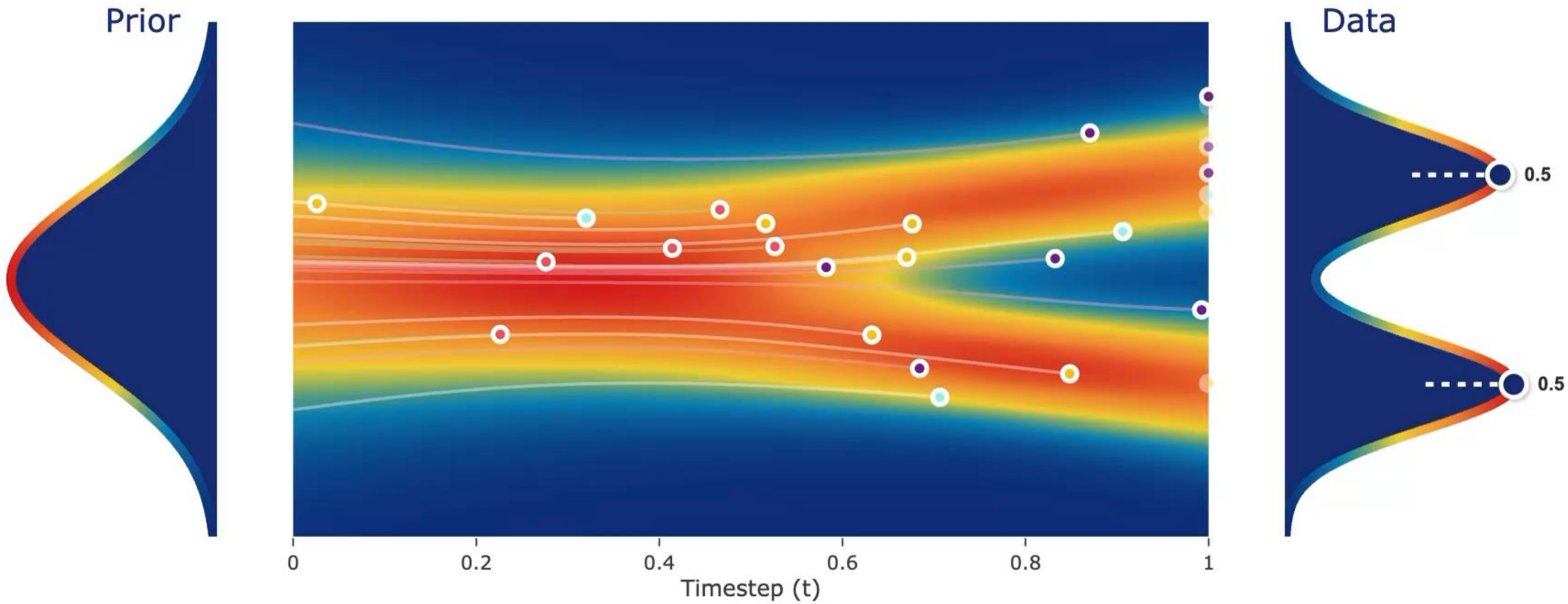
What will neural net predict? (Marginal flow)

$$L = ||(x_{\text{clean}} - x_{\text{noise}}) - u_\theta(x_t)||^2$$



Don't forget: $x_t = tx_{\text{clean}} + (1-t)x_{\text{noise}}$

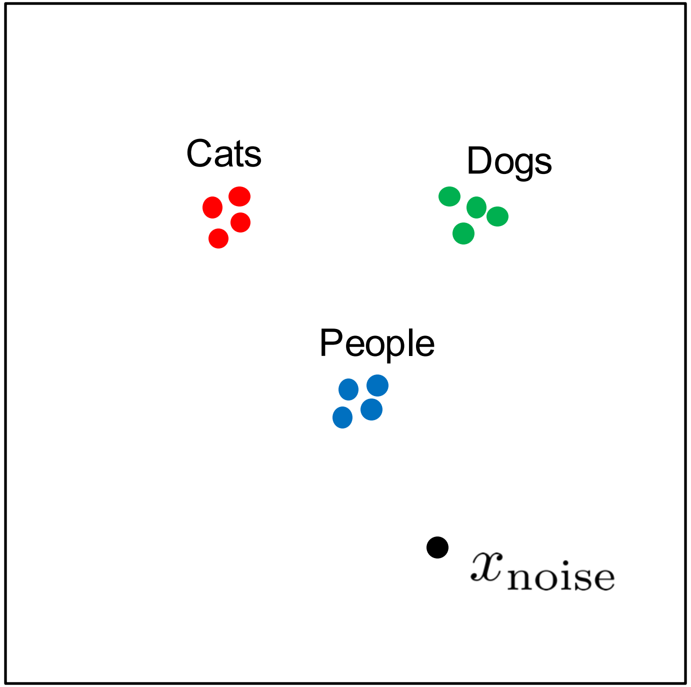
*As t moves closer to clean, it is biased more to a real data point.



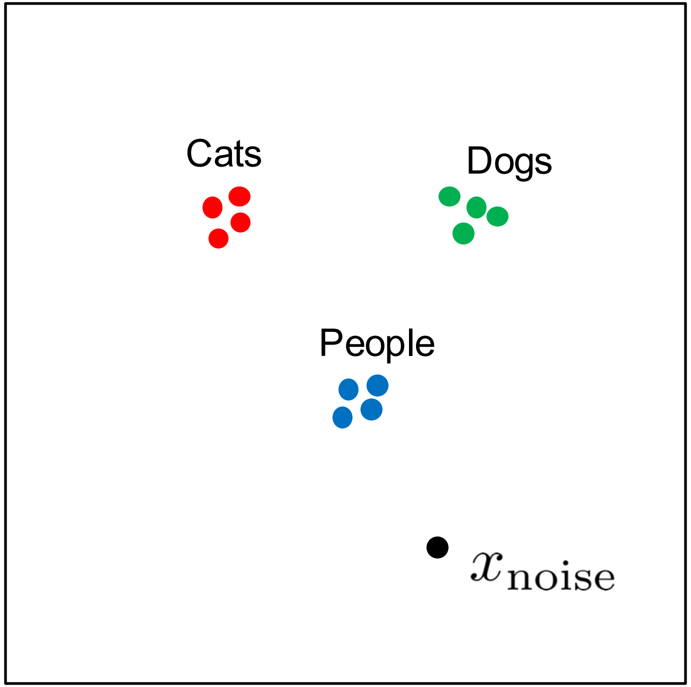
McAllister 2025

Classifier-free guidance (CFG)

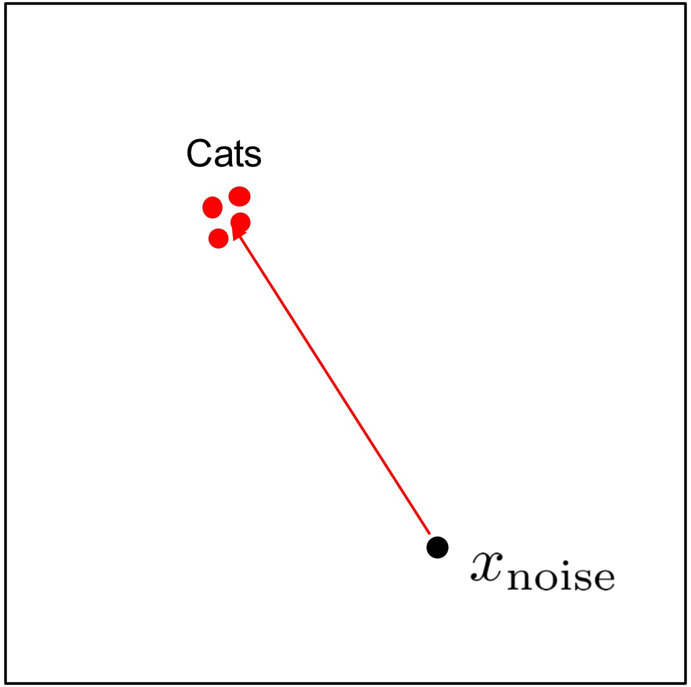
Now, imagine 2D world with 3 types of points



I want a “Cats” model, what’s the marginal flow?

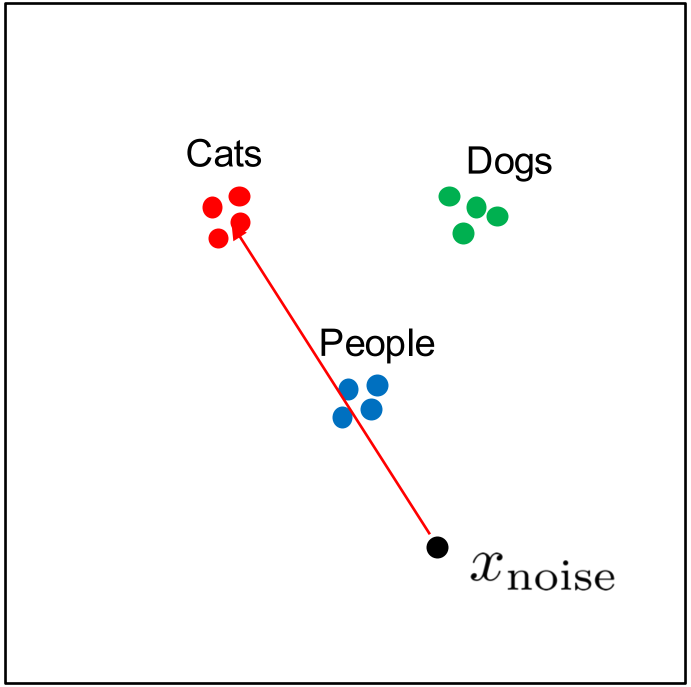


I want a “Cats” model, what’s the marginal flow?



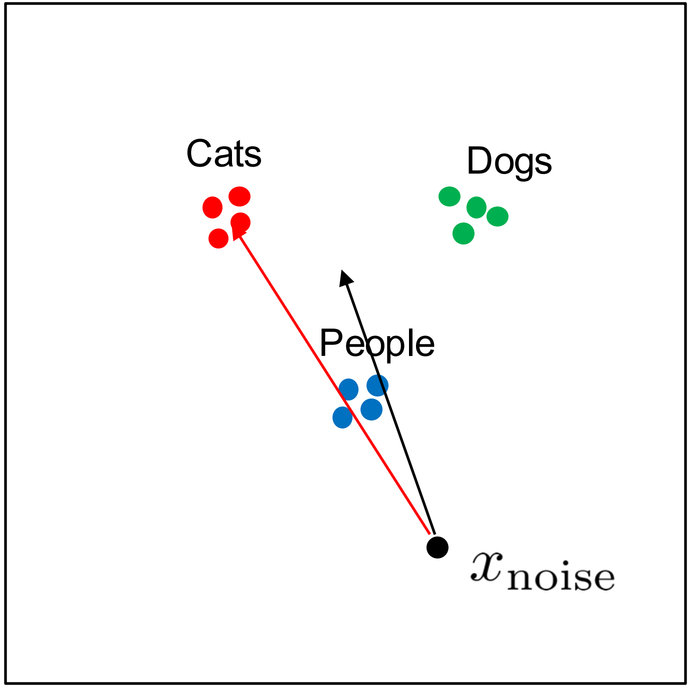
*At noise, point to the average!

What's unconditional marginal flow?



*At noise, point to the average!

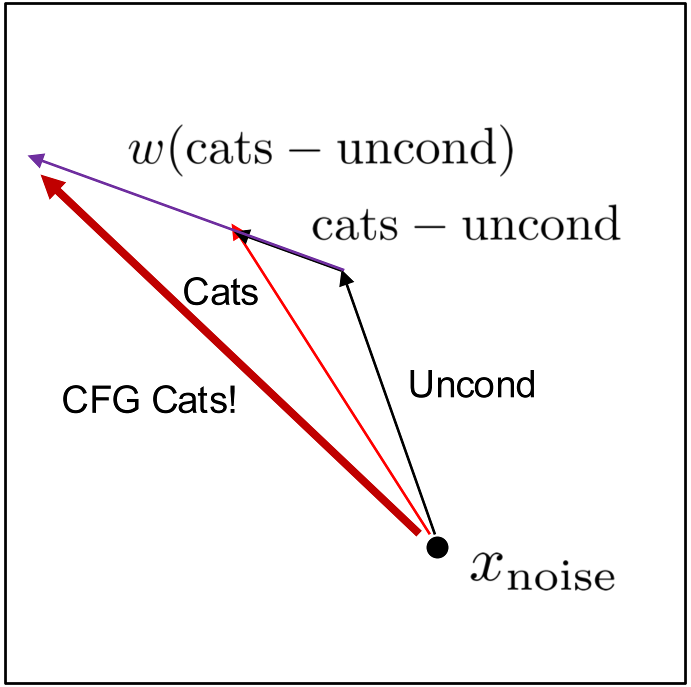
What's unconditional marginal flow?



*At noise, point to the average!

CFG = Exaggerate the "class" more

$$\text{CFG} = \text{uncond} + w(\text{cats} - \text{uncond})$$



Let's do worksheets!

*Why? People like catty-images not cat-ish images