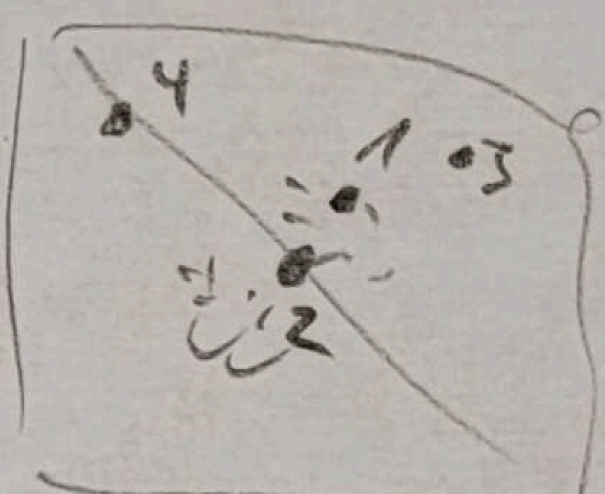
12/8/21

1/

① Slide w/ Alex' classification
w/o classif. prob

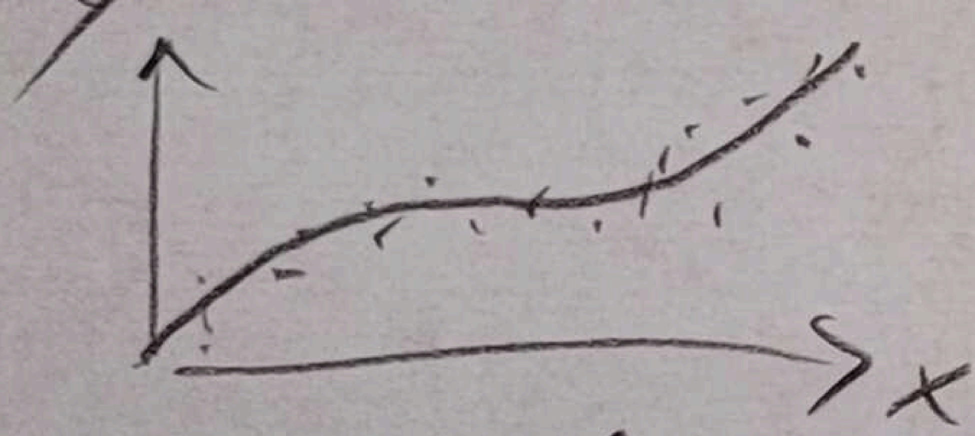


→ add 4 test dots

→ uncertainty for OOD samples

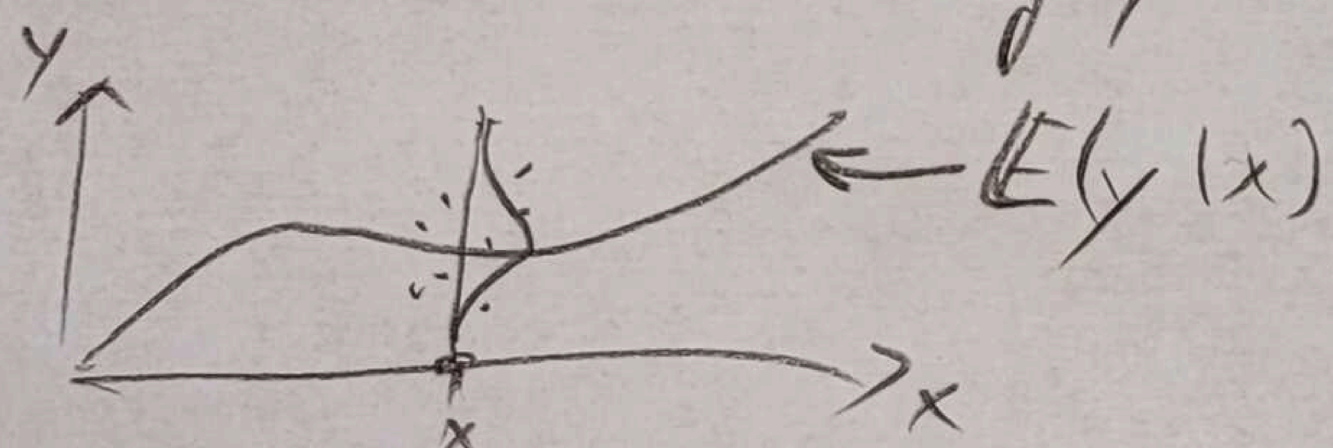
② Modeling uncertainty (regression, 1D)

Ⓐ Vanilla



$y = f(x)$

Ⓑ Prob. treatment of y



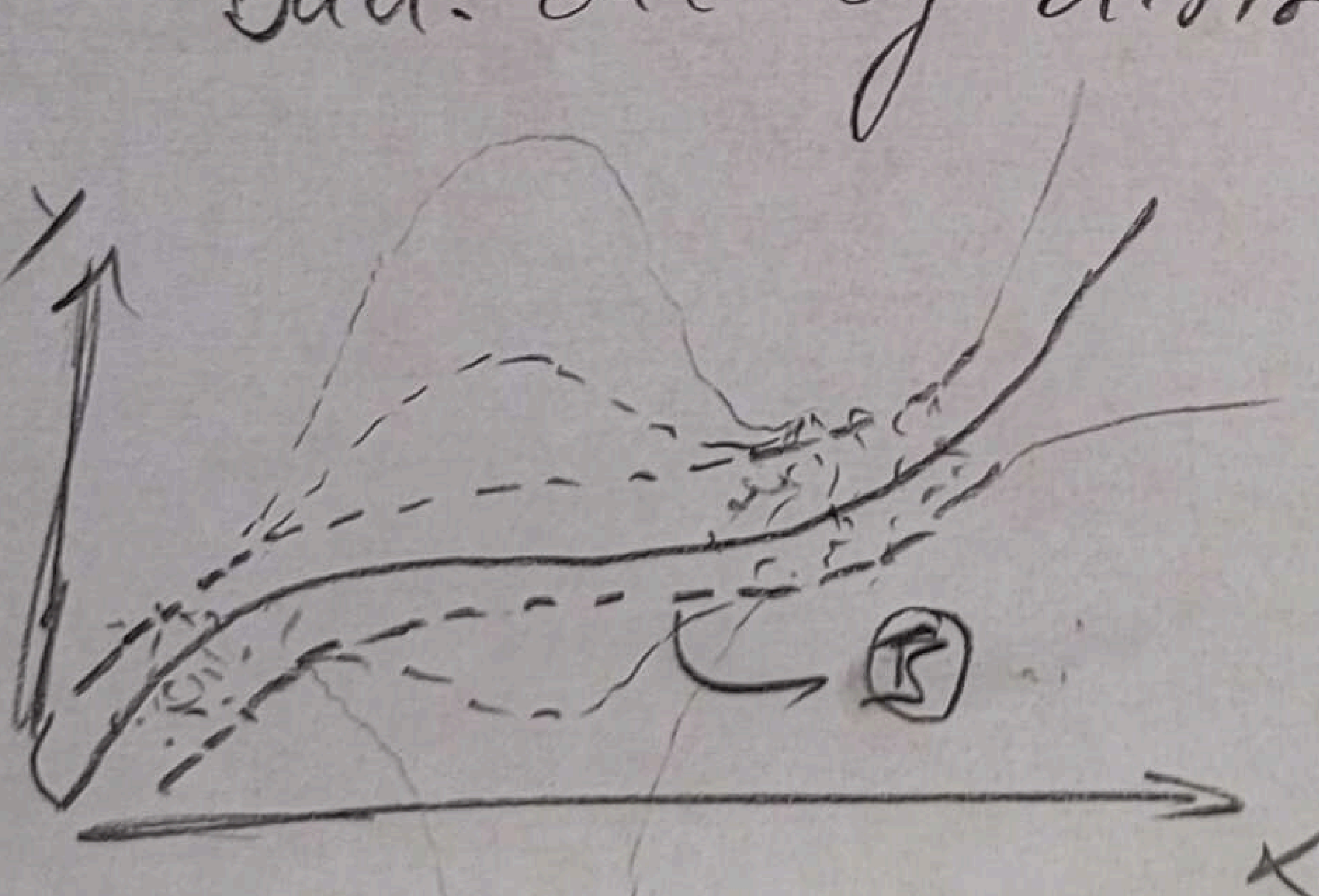
$y \sim N(f(x); \sigma^2)$

→ good: captures uncertainty "close to training data".

[remark: does not imply that y is truly stochastic; e.g. other variables]

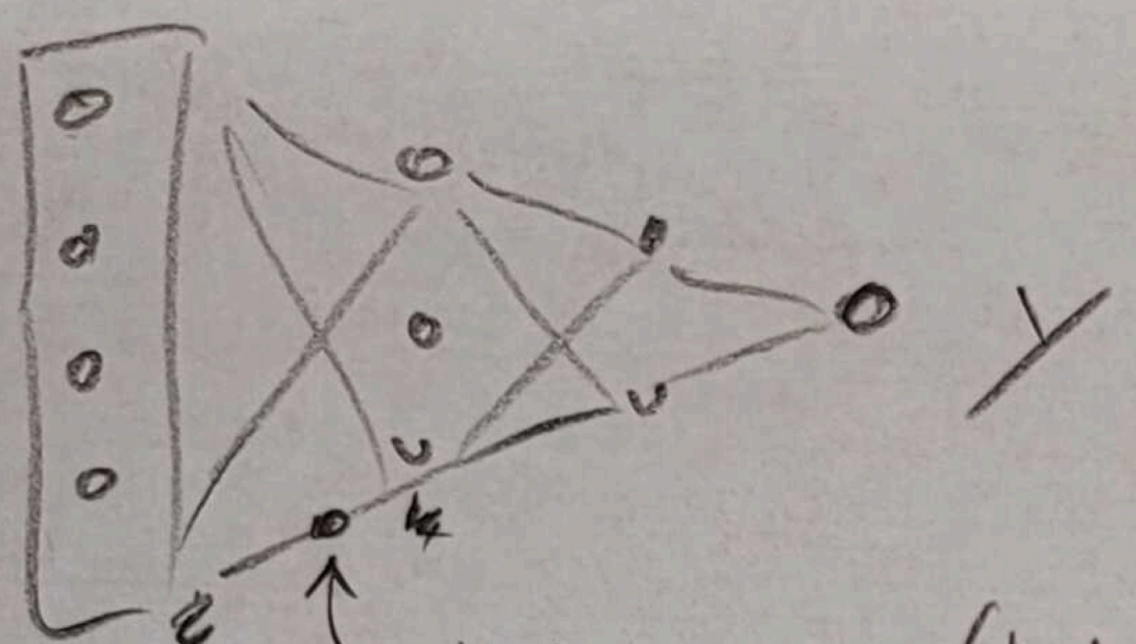
bad: out-of-distribution

Ⓒ



Solutions?

- Gaussian process?
  → nice, but slow if many training samples
- generally a fct.
  $f(x, \theta)$ w/ $p(\theta)$
- BNN is such a fct.



X

$f(x; w)$

$W \sim p(W \mid \text{training data})$

↳ Generally:

Prior: $p(W)$ e.g. $= \prod_{d=1}^{D} \mathcal{N}(W_i; 0, 1)$

$X, Y = (x^{(1)}, x^{(2)}, \ldots^{(N)}), (y^{(1)}, y^{(2)}, \ldots)$

Likelihood:

$p(Y \mid X, W) = \prod_{i=1}^{N} \mathcal{N}(y^{(i)}; f(x^{(i)}; W), \sigma^2)$

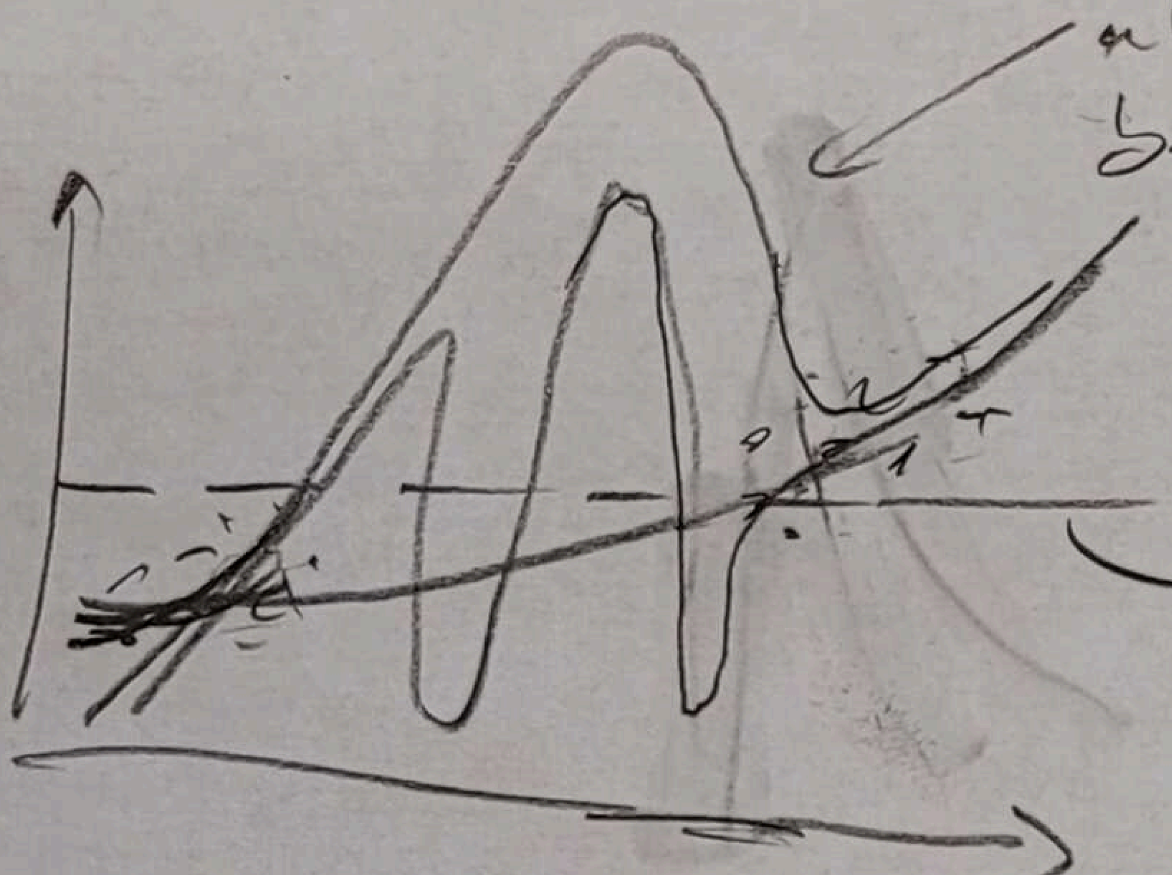
Bayes rule:

$$p(W \mid X, Y) = \frac{p(W) \cdot \prod_i \mathcal{N}(y^{(i)}; f(x^{(i)}; W), \sigma^2)}{\text{Norm.}}$$

① This is hard to calculate!

② There can be complicated dependencies between $p(W_i, W_j \mid X, Y) \ldots$

③ But, if we can solve this, what do we gain?



all valid models! but the "wild" ones will have large $|W|$ → smaller probability b/c prior

also valid, but unlikely b/c of data likelihood

"Average-y" (marginalization)

$$\bar{y} = \int p(W \mid \text{training})\, f(x, W)\, dW$$

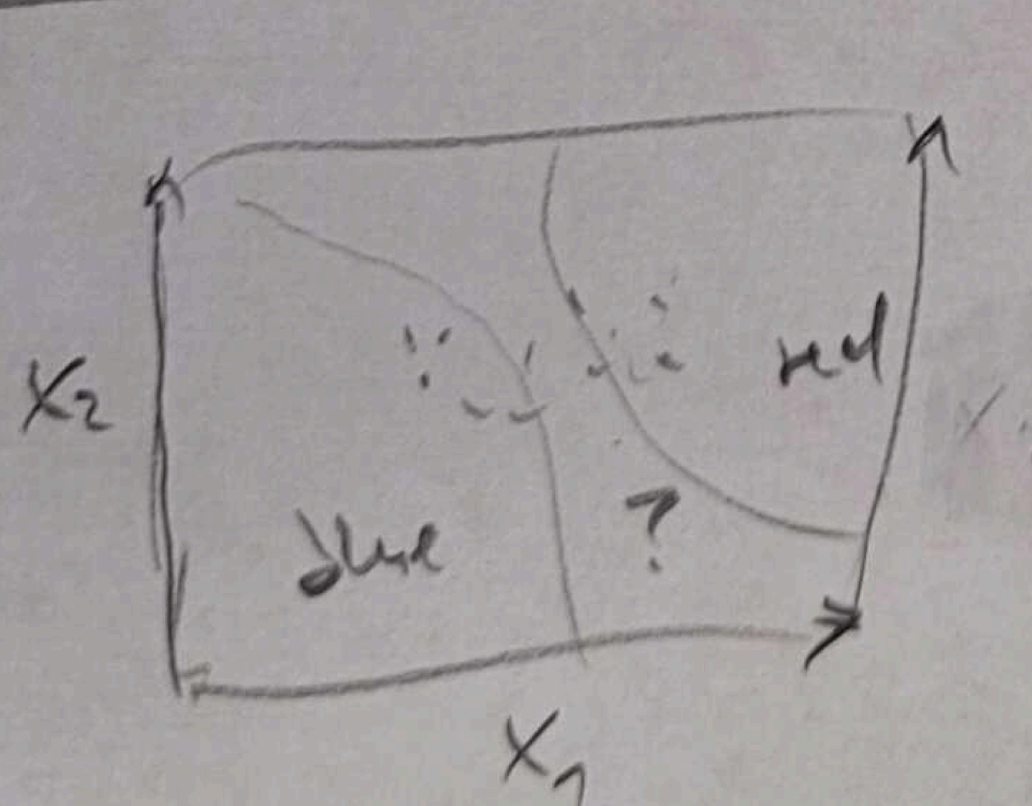
↳ integral is hard if $p(W \mid \text{data})$ and f is complicated.

↳ sampling

$$\boxed{\bar{y} \approx \frac{1}{S} \sum_{s=1}^{S} f(x, W^{(s)}) \text{ w/ } W^{(s)} \sim p(W \mid \text{data})}$$

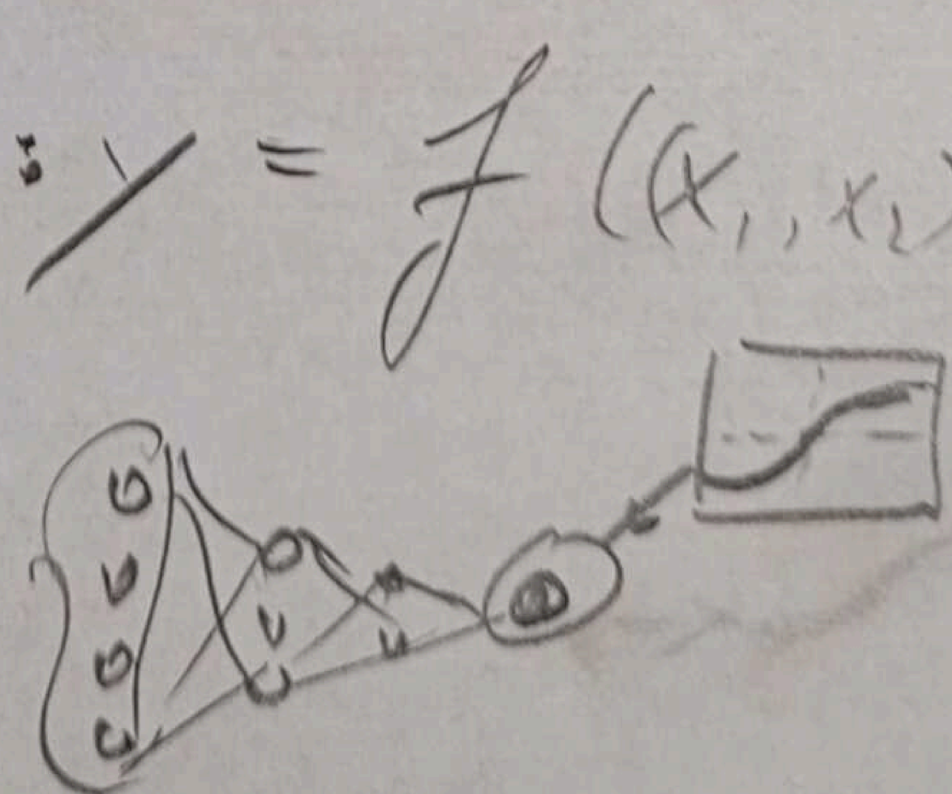
likewise $\mathbb{E}[(y-\bar{y})^2] = \text{var}\, y$

Link back to classification:



$$p(\text{red}) =: \underset{\in [0,1]}{y} = f((x_1, x_2), W)$$

w/ f =



Classification game

...