LECTURE 1 : LINEAR REGRESSION

let $\{x_i, y_i\}_{i=1,..,n}$ a set of input - output pairs

the goal of a regression analysis is to find a model

$$\hat{y} = f(\underbrace{x}_{\text{observation}}, \overbrace{w}^{\text{parameter of the model}})$$

that given a new observation x_{n+1}

it predicts the output y_{n+1} as closely as possible.

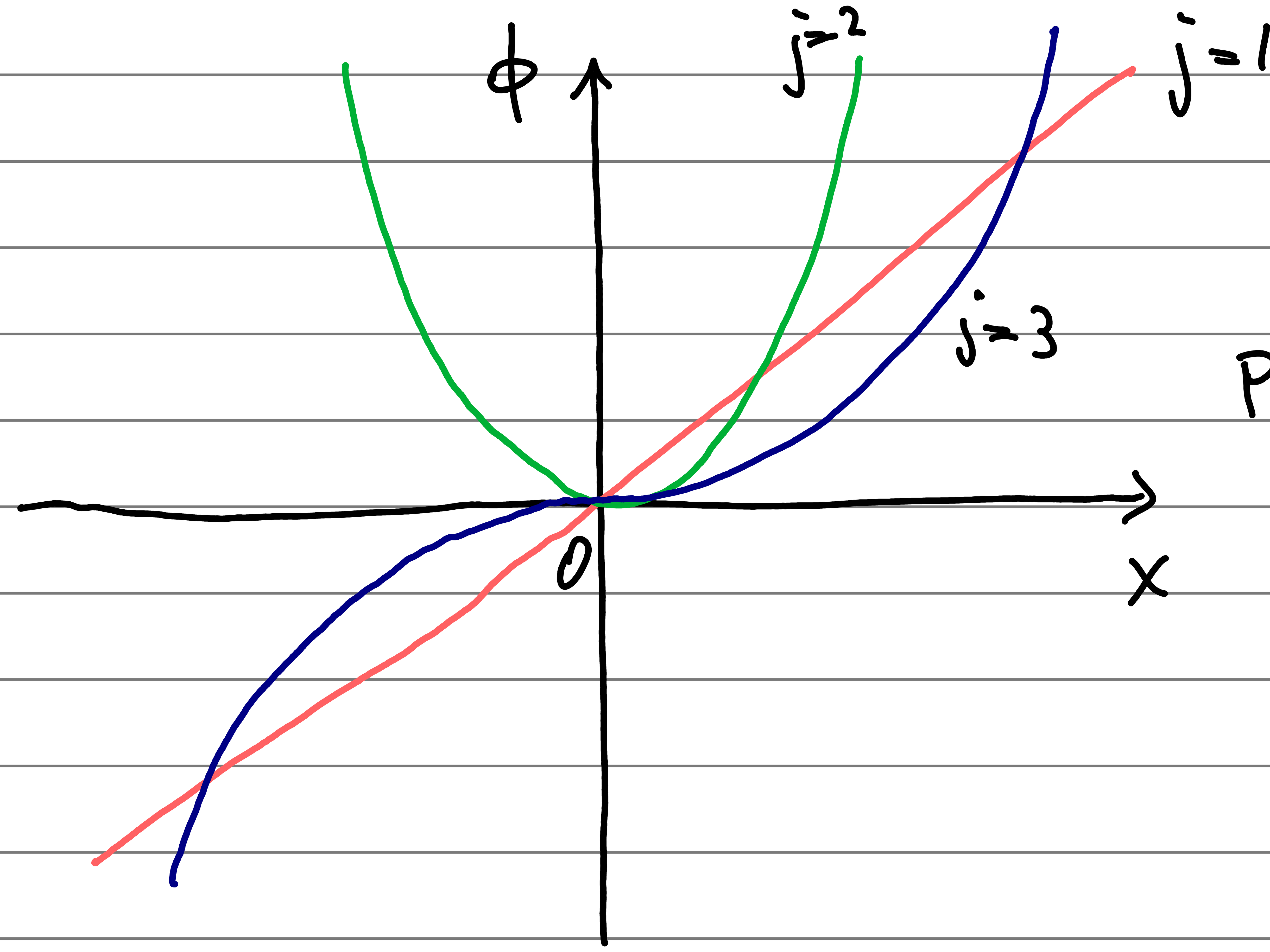
→ LINEAR regression refers to the nature of the model $\hat{y}$

$$\hat{y} = f(\underline{x}, \underline{w}) = \underbrace{w_0}_{\text{bias}} + \sum_{j=1}^{M-1} \underbrace{w_j}_{\text{weights}} \underbrace{\phi_j(\underline{x})}_{\text{basis function}}$$

the LINEARITY of f is in the parameters $\{w_i\}_{i=1,..,M}$

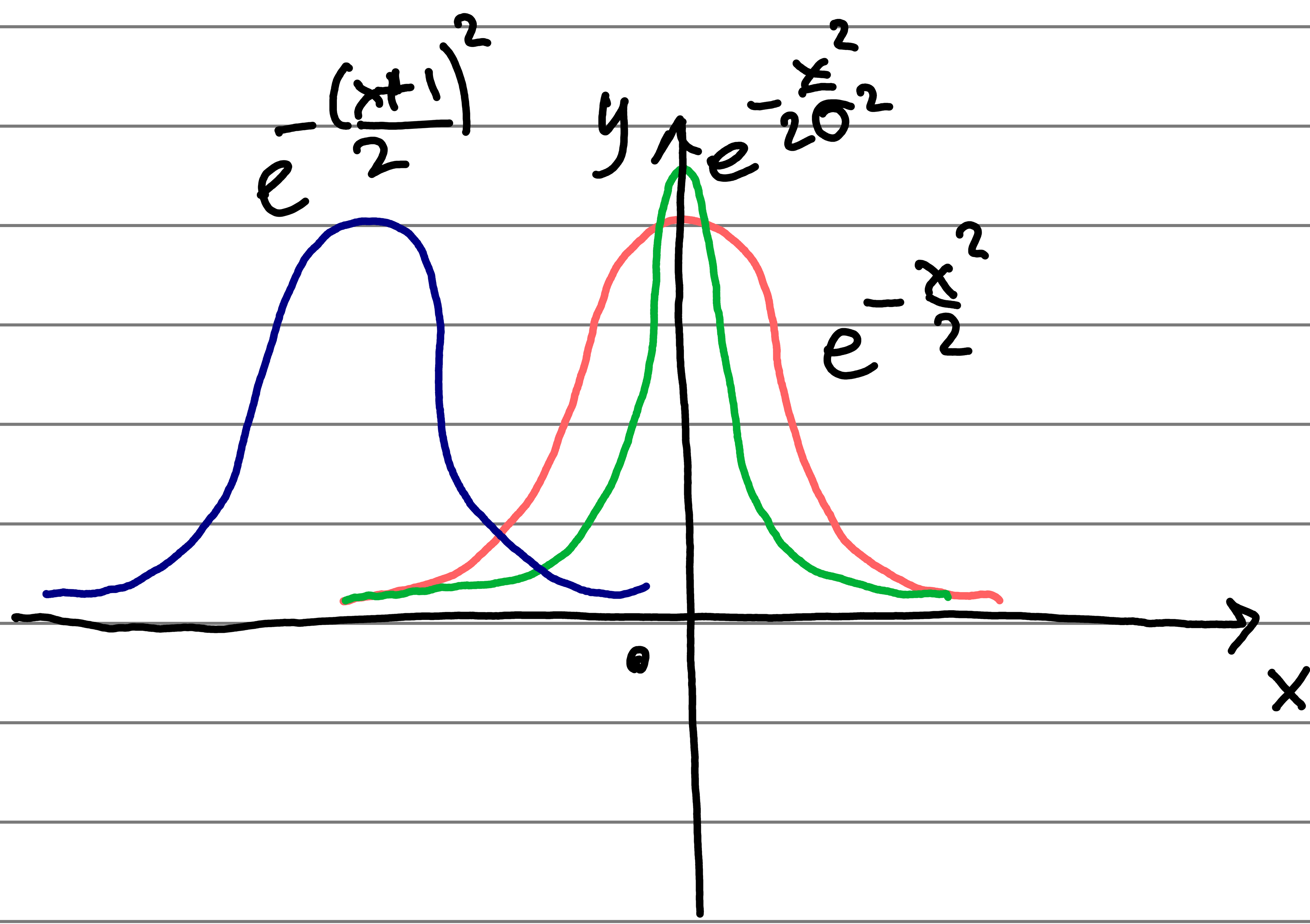
while $\{\phi_j\}_{j=1,..,M}$ can be arbitrary functions

EXAMPLES of BASIS FUNCTIONS



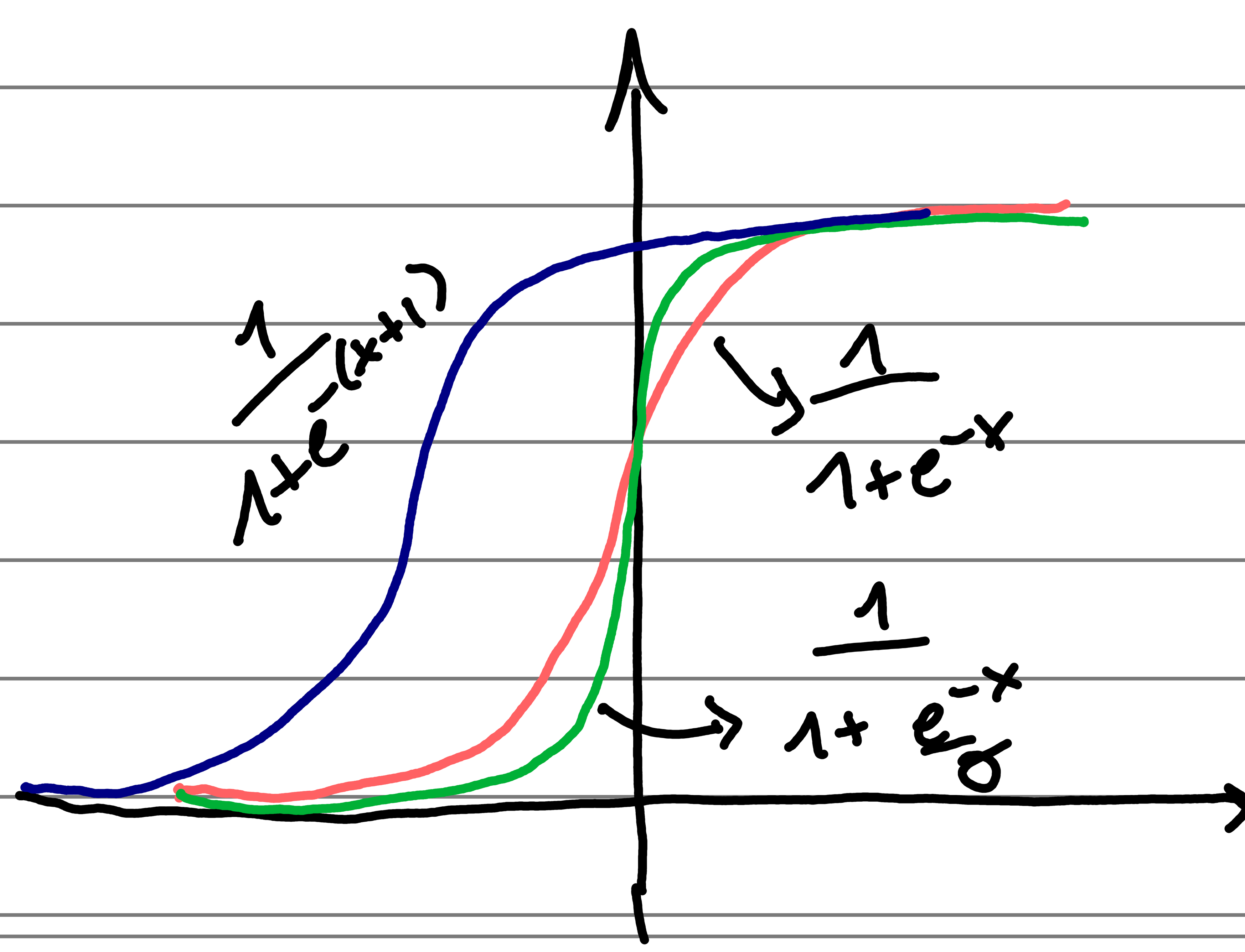
$$\phi_j = x^j$$

POLYNOMIALS



$$\phi_j = e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}$$

GAUSSIANS



$$\phi_j = \frac{1}{1+e^{-\frac{(x-\mu_j)}{\sigma_j}}}$$

MAXIMUM LIKELIHOOD AND LEAST SQUARES

how do we get the best model $\hat{y}$?

$$\hat{y} = f(\underline{x}, \underline{w}) + \epsilon$$

↳ Gaussian additive noise

$$p(\epsilon) = \mathcal{N}\left(\mu = 0, \sigma = \frac{1}{\beta}\right)$$

↳ normal dist.

let's write down the probability of getting an output y given a model $\hat{y}$

$$p(\underline{y} \mid \underline{\hat{y}}) \equiv p(\underline{y} \mid \underline{x}, \underline{w}, \beta)$$

$$= \mathcal{N}\left(\underline{y} \mid f(\underline{x}, \underline{w}), \frac{1}{\beta}\right)$$

the probability is a Gaussian with mean

$\mu = f(\underline{x}, \underline{w})$ ← the prediction and variance $\frac{1}{\beta}$

$\boxed{\text{FROM NOW ON}}$

- we drop the dependence of p on $\underline{x}$
- we write f in condensed form $f(\underline{x}, \underline{w}) = \underline{w}^T \underline{\phi}(\underline{x})$

if all data taken are <u>independent</u>

$$p(\underline{y}\,|\,\underline{w},\beta) = \prod_{i=1}^{m} \mathcal{N}\left(y_i \,|\, \underline{w}^T\phi(x_i), \frac{1}{\beta}\right)$$

now let's take the logarithm →
- multiplication ↳ addition
- numerical stability
- no exponential

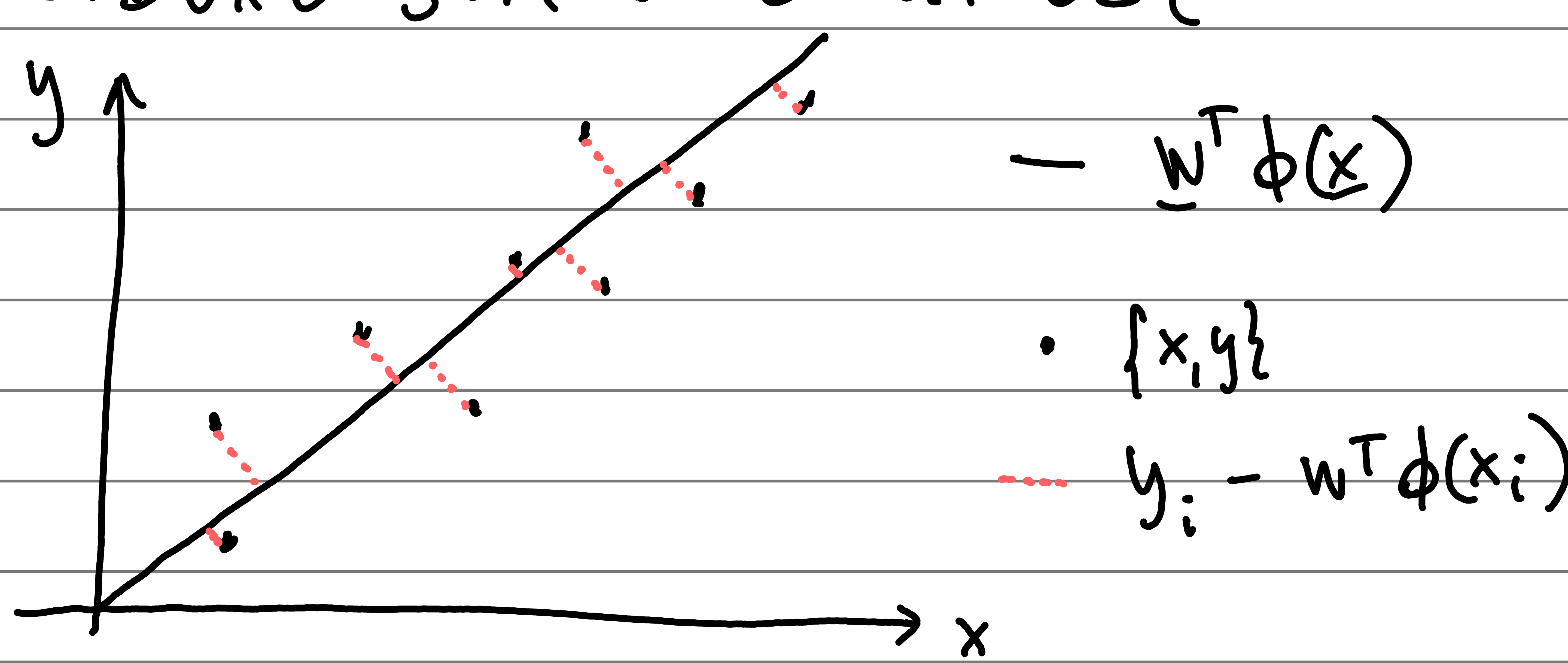
$$\ln\left[p(\underline{y}\,|\,\underline{w},\beta)\right] = \sum_{i=1}^{m} \ln\left[\mathcal{N}\left(y_i \,|\, \underline{w}^T\phi(x_i), \frac{1}{\beta}\right)\right]$$

$$= \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) - \beta E(\underline{w})$$

$$\boxed{E(\underline{w}) = \frac{1}{2}\sum_{i=1}^{m}\left[y_i - \underline{w}^T\phi(x_i)\right]^2}$$

RESIDUAL SUM OF SQUARES (or ERROR FUNCTION)

visually:



goal: maximise $p(\underline{y} \mid \underline{w}, \beta)$

$$\Downarrow$$

minimise $E(\underline{w})$

$$\nabla_W \ln \left[p(\underline{y} \mid \underline{w}, \beta) \right] = \sum_{i=1}^{n} \left[y_i - \underline{w}^T \phi(x_i) \right] \phi^T(x_i)$$

$$\overset{!}{=} 0 \quad \text{(minimise)}$$

solution for $\underline{w}$ gives

$$\underline{w} = \left(\Phi^T \Phi \right)^{-1} \Phi^T \underline{y}$$

where
$$\Phi = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \dots & \phi_{M-1}(x_1) \\ \vdots & & & \\ \vdots & & & \\ \phi_0(x_M) & \phi_1(x_M) & \dots & \phi_{M-1}(x_M) \end{bmatrix} \quad N \times M \text{ matrix}$$

$$\Phi^\dagger = \left(\Phi^T \Phi \right)^{-1} \Phi^T \quad \text{Moore-Penrose pseudoinverse}$$