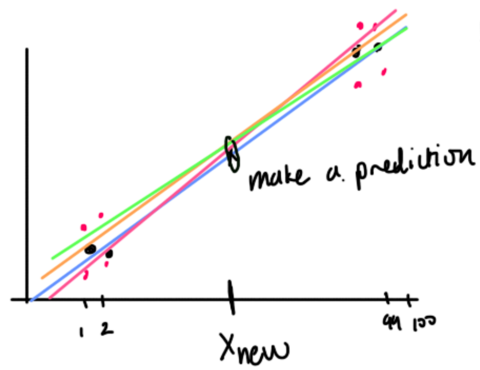
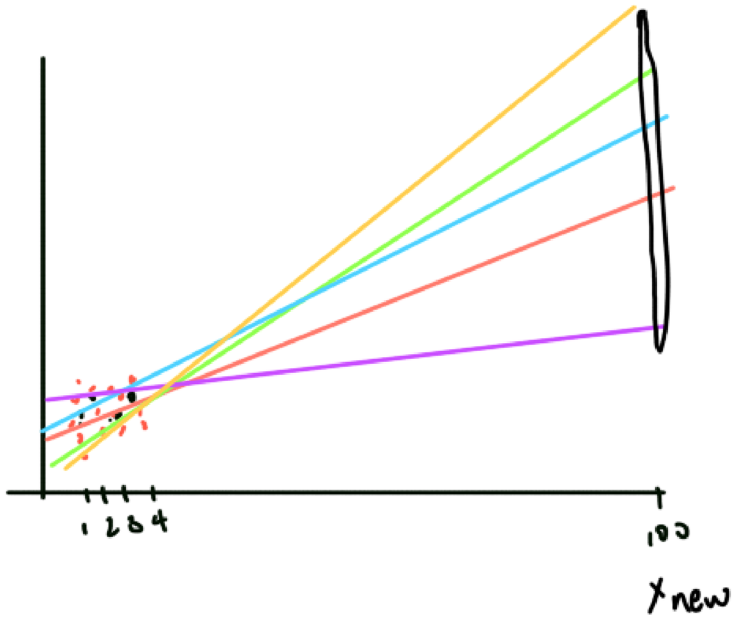


fit a linear model

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 99 \\ 1 & 100 \end{bmatrix}$$

$x_{new} = 50$

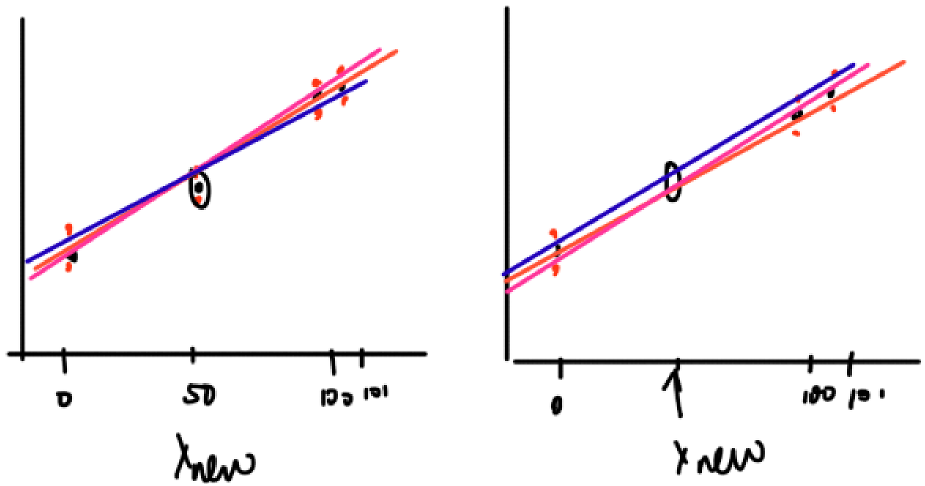
↑ not much variation for the prediction for x_{new}.

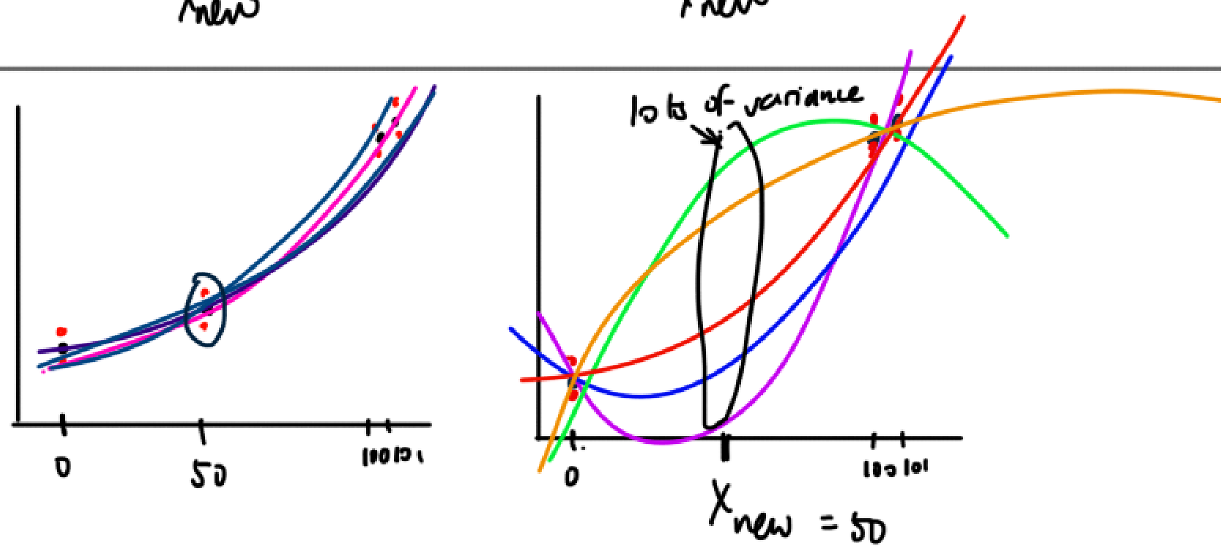


$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

$x_{new} = 100$

very high variation in the predictions


0
50
100
101
Xnew
0
100
101
Xnew


0
50
100 101
lots of variance
0
100 101
Xnew = 50

Multivariate normal distribution

First, the univariate normal distribution: $X \sim N(\mu, \sigma^2)$

density: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

The multivariate normal: $\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$ $\quad \vec{X} \sim N(\vec{\mu}, \Sigma)$

$\vec{X}$ is a d dimensional vector

$\vec{\mu}$ = d-vector of means

Σ = $d \times d$ positive definite variance matrix

density $f(\vec{x}) = f(x_1, x_2, \ldots x_d) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right]$

Some similarities between m.v. normal and univariate

$(\vec{x}-\vec{\mu})^T(\vec{x}-\vec{\mu})$ is similar to $(x-\mu)^2$

Σ^{-1} is similar to $\frac{1}{\sigma^2}$

$|\Sigma|$ is the determinant of Σ.

The determinant is equal to the area or volume of the parallelogram/parallelepiped of the vectors in Σ

Its role in the PDF is to ensure the PDF integrates to 1.

Example 1:

$d = 3 \qquad \vec{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad \Sigma = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad |\Sigma| = 1 \qquad \Sigma^{-1} = I$

$$f(\vec{x}) = f(x_1, x_2, x_3) = \left(\frac{1}{\sqrt{(2\pi)^3}} \right) \exp\left(-\frac{1}{2} \vec{X}^T I \vec{X} \right)$$

$$= \left(\frac{1}{\sqrt{2\pi}} \right)^3 \exp\left(-\frac{1}{2} \left(x_1^2 + x_2^2 + x_3^2 \right) \right)$$

$$= \underbrace{\left(\frac{1}{\sqrt{2\pi}} \exp -\frac{x_1^2}{2} \right)}_{X_1 \sim N(0,1)} \underbrace{\left(\frac{1}{\sqrt{2\pi}} \exp \frac{-x_2^2}{2} \right)}_{X_2 \sim N(0,1)} \underbrace{\left(\frac{1}{\sqrt{2\pi}} \exp \frac{-x_3^2}{2} \right)}_{X_3 \sim N(0,1)}$$

$f(\vec{x})$ is the product of three iid $N(0,1)$

for any d and diagonal matrix Σ (all non diagonal entries are 0)

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{bmatrix} \qquad |\Sigma| = \prod_{i=1}^{d} \sigma_i^2 \qquad \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ 0 & \frac{1}{\sigma_2^2} & & \\ \vdots & & \ddots & \\ & & & \frac{1}{\sigma_d^2} \end{bmatrix}$$

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(\frac{-1}{2} \frac{(x_1-\mu_1)^2}{\sigma_1^2}\right) \cdot \cdots \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\left(\frac{-1}{2} \frac{(x_d-\mu_d)^2}{\sigma_d^2}\right)$$

$f(\vec{x})$ is a product of the PDFs of d normal distributions,

each $X_i \sim N(\mu_i, \sigma_i^2)$

Example: $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_3\left(\underbrace{\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}}_{\vec{\mu}}, \underbrace{\begin{bmatrix} 4 & 2 & 0 \\ 2 & 1 & -1 \\ 0 & -1 & 9 \end{bmatrix}}_{\Sigma} \right)$

the marginal distributions are:

$X_1 \sim N(\underbrace{1}_{\mu_1}, \underbrace{4}_{\sigma_1^2})$ $\qquad X_2 \sim N(-1, 1)$ $\qquad X_3 \sim N(2, 9)$

$cov(X_1, X_2) = \sigma_{12} = \sigma_{21} = 2$

$cov(X_1, X_3) = \sigma_{13} = \sigma_{31} = 0 \Rightarrow X_1 \perp X_3$

$cov(X_2, X_3) = \sigma_{23} = \sigma_{32} = -1$

Example: $\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_3\left(\underbrace{\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}}_{\vec{\mu}}, \underbrace{\begin{bmatrix} 4 & 2 & 0 \\ 2 & 1 & -1 \\ 0 & -1 & 9 \end{bmatrix}}_{\Sigma} \right)$

subsets of $\vec{X}$ also follow the multivariate normal:

$\vec{X}_K = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ then $\mu_K = \begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix}$ and $\Sigma_K = \begin{bmatrix} \sigma_1^2 & \sigma_{13} \\ \sigma_{31} & \sigma_3^2 \end{bmatrix}$

$\begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} \right)$

If $\vec{X} \sim N(\vec{\mu}, \Sigma)$ and C is a $d \times d$ non-singular matrix,

then $\vec{Y} = C\vec{X} \sim N(C\vec{\mu}, C\Sigma C^T)$

example: $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \right)$

let $\vec{Y} = \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix}$

$\vec{Y}$ is a linear combination of the variables in $\vec{X}$, which can be expressed with a matrix C.

$C = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

$C\vec{X} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix} = \vec{Y}$

$$\begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right)$$

$$\sim N\left(\begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} \right)$$

$$\vec{Y} = C\,\vec{X}$$

$$E(\vec{Y}) = E(C\vec{X}) = C\,E(\vec{X}) = C\vec{\mu}$$

$$\mathrm{Var}(\vec{Y}) = E\left((C\vec{X} - C\vec{\mu})(C\vec{X} - C\vec{\mu})^T\right)$$

$$= E\left(C\,(\vec{x}-\vec{\mu})\,(\vec{x}-\vec{\mu})^T C^T \right.$$

$$= C\;E\left((\vec{x}-\vec{\mu})\,(\vec{x}-\vec{\mu})^T\right)\,C^T$$

$$= C\;\Sigma\;C^T$$