# Lecture 9

Monday, September 25, 2023 10:04 AM

## Bivariate

$$\begin{array}{cccc} \vec{1} & X & Y & \hat{Y} \\ \begin{bmatrix}1\\1\\1\\\vdots\\1\end{bmatrix} & \begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} & \begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix} & \begin{bmatrix}\hat{y}_1\\\hat{y}_2\\\vdots\\\hat{y}_n\end{bmatrix} \end{array}$$

Independent Variable   target   predicted

$$\hat{y}_i = w x_i + b \quad \forall\, i = 1..n$$

$$\hat{y}_i = f(x_i)$$

Unknown: $w$ : slope

$b$ : bias

$$\boxed{\hat{Y} = w.X + b.\vec{1}}$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Sum of squared error

$$\boxed{\frac{\partial SSE}{\partial w} = 0} \qquad \boxed{\frac{\partial SSE}{\partial b} = 0}$$

$$w = \frac{Cov(X,Y)}{Var(X)} \qquad \boxed{b = \mu_Y - w.\mu_X}$$

---

$\vec{1}, X \longrightarrow \hat{Y}$   $Y$

input   prediction   true

$\in \mathbb{R}^n$

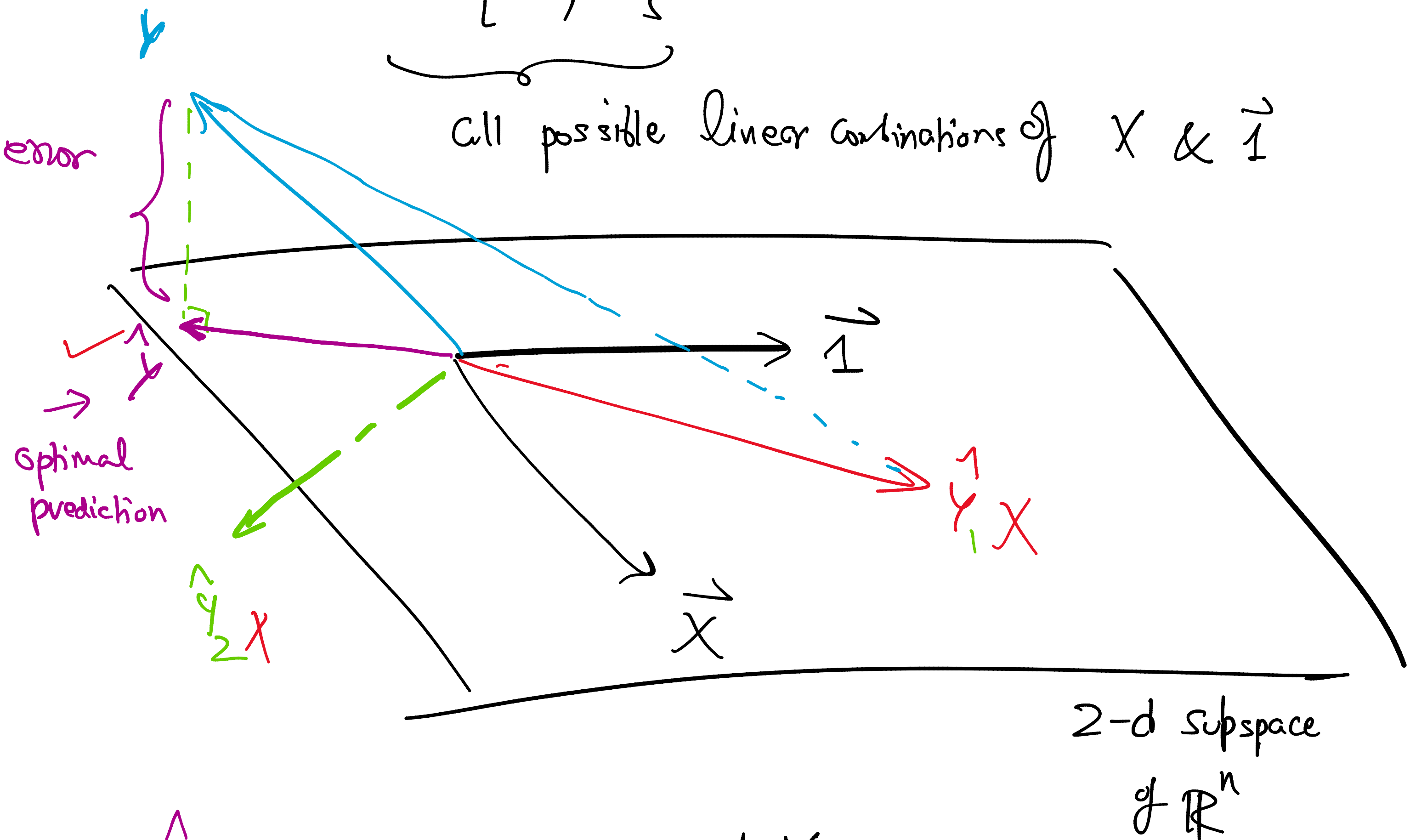
n-dim space

$$\hat{Y} = w.X + b.\vec{1}$$

$$Y = w \cdot \underline{X} + b \cdot \underline{\bar{1}}$$

n-dim space
"Col-view"

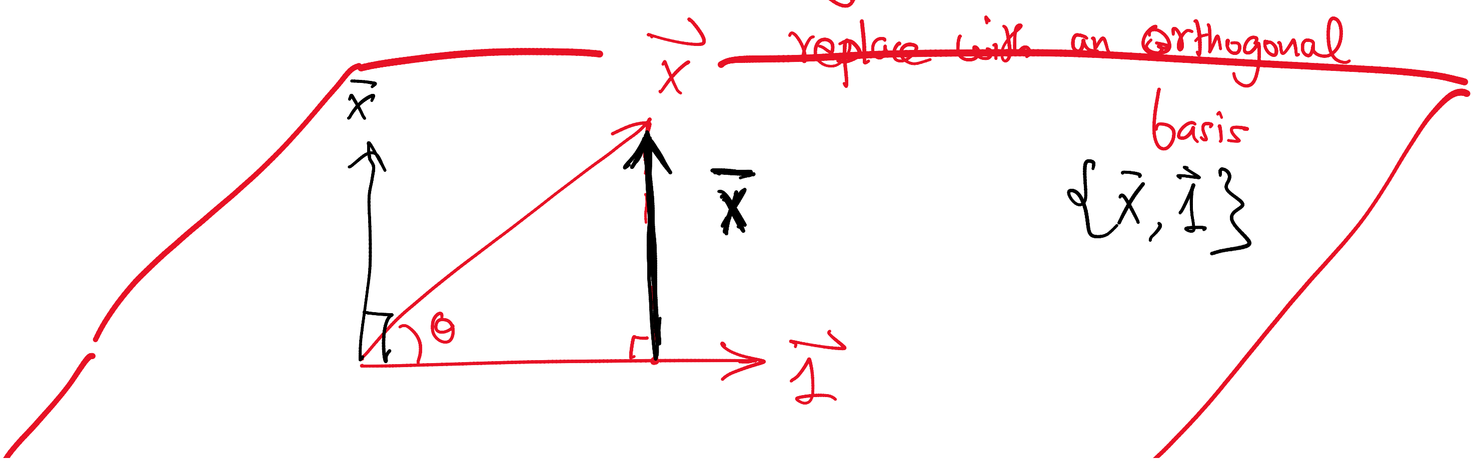
$$\hat{Y} \in \underbrace{\text{span}\{X, \vec{1}\}}_{\text{all possible linear combinations of } X \,\&\, \vec{1}}$$




$\hat{Y}$ = Orthogonal projection of $Y$ onto span$\{\vec{X}, \vec{1}\}$ ⇐ original basis

⇓

replace with an orthogonal basis $\{\bar{x}, \vec{1}\}$



$\bar{X}, \vec{1}$ are orthogonal

$$\bar{X} = X - \text{proj}_{\vec{1}}(X) \cdot \vec{1}$$

orthogonalization

$$= X - \left(\frac{X^T 1}{1^T 1}\right)\vec{1}$$

$$\frac{X^T 1}{1^T 1} = \frac{\sum x_i}{n} = \mu_x$$

$$\bar{X} = X - \mu_x \cdot \vec{1}$$

$\hat{Y}$ = project $Y$ onto span $\{\bar{X}, \vec{1}\}$ orthogonal basis

$$\hat{Y} = \text{Proj}_{\bar{X}}(Y) \cdot \bar{X} + \text{Proj}_{\vec{1}}(Y) \cdot \vec{1}$$

$$\hat{Y} = \underbrace{\left(\frac{Y^T \bar{X}}{\bar{X}^T \bar{X}}\right)}_{\omega} \bar{X} + \underbrace{\left(\frac{Y^T 1}{1^T 1}\right)}_{\mu_Y} \vec{1}$$

$$\hat{Y} = \omega \cdot \bar{X} + \mu_Y \cdot 1$$

$$=$$

$$= w\left(X - \mu_X \cdot \vec{1}\right) + \mu_Y \cdot \vec{1}$$

$$= wX + \underbrace{\left(\mu_Y - w\mu_X\right)}_{b} 1$$

$$= w \cdot X + b \cdot \vec{1}$$

$$b = \mu_Y - w \cdot \mu_X$$

$$w = \frac{\bar{Y}^T \bar{X}}{\bar{X}^T \bar{X}} = \frac{\mathrm{cov}(X,Y)}{\mathrm{ver}(X)}$$

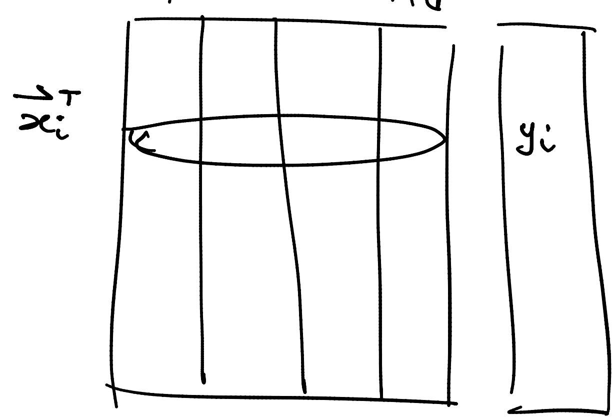
---

d-dim

$A_i \equiv X_i$

Multiple regression

Independent vars → $X_1 \ X_2 \ \ldots \ X_d$ $\quad Y$ ← target varible $\mathbb{R}$




$$\hat{y}_i = f(\vec{x}_i)$$

$\vec{x}_i \in \mathbb{R}^d \qquad y_i \in \mathbb{R}$

real value

Multivariate regression

$X_1 X_2 \ldots X_d \quad Y_1 Y_2 \ldots Y_q$

$$\hat{\vec{y}}_i = f(\vec{x}_i)$$

real value

$$y_i = f(\vec{x}_i)$$
$$\hat{y}_i \in \mathbb{R}^q$$
$$\vec{x}_i \in \mathbb{R}^d$$

## linear model

$$\boxed{\hat{y}_i = w_1 x_{i1} + w_2 x_{i2} + \cdots + w_d x_{id} + b}$$
$$= f(\vec{x}_i)$$

$$\vec{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix}$$

$$\boxed{\hat{y}_i = \vec{w}^T \vec{x}_i + b}$$

scalar

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$$

$\vec{w} \equiv$ weight vector

$b \equiv$ bias

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\boxed{SSE = \sum_{i=1}^{n} (y_i - w^T \vec{x}_i - b)^2}$$

$$\underbrace{\frac{\partial SSE}{\partial \vec{w}} = 0 \qquad \frac{\partial SSE}{\partial b} = 0}_{\text{Solve!}}$$

n - equations

$$\hat{y}_1 = w_1 x_{11} + w_2 x_{12} + \cdots + w_d x_{1d} + b \cdot 1$$
$$\hat{y}_2 = w_1 x_{21} + \ldots$$

$$\hat{y}_1 = w_1 x_{11} + w_2 x_{12} + \cdots + w_d x_{1d} + b \cdot 1$$

$$\hat{y}_2 = w_1 x_{21} + w_2 x_{22} + \cdots + w_d x_{2d} + b \cdot 1$$

$$\vdots$$

$$\hat{y}_n = w_1 x_{n1} + w_2 x_{n2} + \cdots + w_d x_{nd} + b \cdot 1$$

$$\hat{Y} \quad \underbrace{X_1 \quad X_2 \quad \cdots \quad X_d}_{D} + \tilde{1}$$

$$\tilde{D} = \text{augmented data}$$

$$\boxed{\hat{Y} = w_1 \cdot X_1 + w_2 \cdot X_2 + \cdots + w_d \cdot X_d + b\,\tilde{1}}$$

$\leftarrow$ 1 vector equation

$$\boxed{\hat{Y} = w_0 \cdot X_0 + w_1 X_1 + w_2 X_2 + \cdots + w_d X_d}$$

$$X_0 = \tilde{1}$$

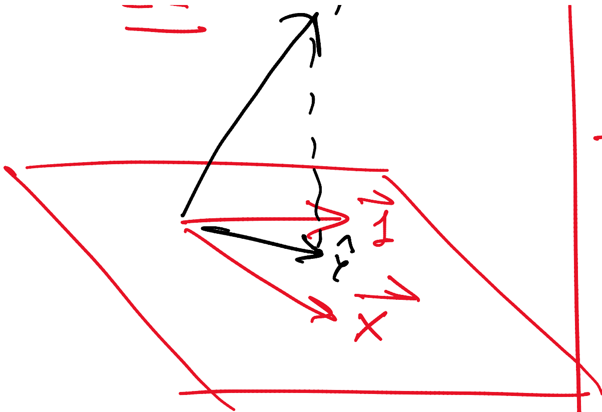
$$\tilde{w} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix} = \begin{pmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \qquad \tilde{D} = n \times (d+1)$$

$(d+1)$

$$\tilde{D} = \begin{bmatrix} | & | & & | \\ X_0 & X_1 & \cdots & X_d \\ | & | & & | \end{bmatrix}$$

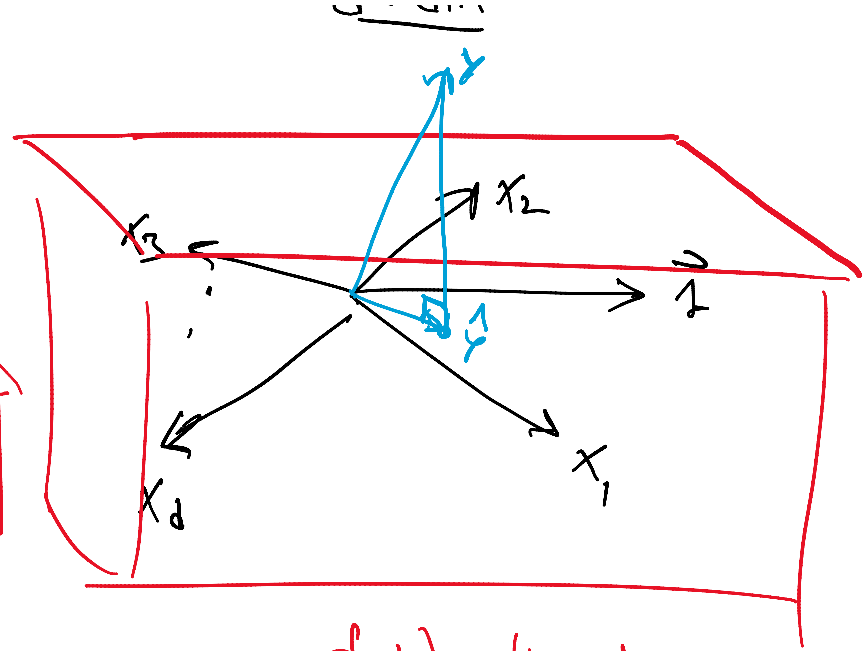
2d

d-dim




2d plane
$\{\vec{X}, \vec{1}\}$

$\mathbb{R}^n$




$(d+1)$ dim subspace of $\mathbb{R}^n$

$d << n$ (typically)

$\hat{Y}$ = orthogonal projection of $Y$ onto span$\{\vec{1}, X_1, X_2, \dots, X_d\}$ Original

$\Downarrow$

Orthogonal basis $\{$ QR decomposition $\}$

a) $\vec{1}$

b) orthogonalize $X_1$ w.r.t $\vec{1}$

c) " $X_2$ w.r.t ortho $X_1$ & $\vec{1}$

orthogonalize $X_i$ with all previous vectors

$\vec{1}$
$\parallel$
$\{U_0, U_1, U_2, \dots, U_d\}$

$\hat{Y} = \text{Proj}_{U_0}(Y) \cdot U_0 +$
$\text{Proj}_{U_1}(Y) \cdot U_1 +$

$$\text{Proj}_{U_1}(Y) \cdot U_1 +$$

$$\vdots$$

$$\text{Proj}_{U_d}(Y) \cdot U_d$$

all previous vectors

---

$$\tilde{D} \equiv \{ \vec{1}, X_1, X_2, \ldots, X_d \}$$

new basis Orthogonal $= \{ U_0, U_1, U_2, \ldots, U_{d-1}, U_d \}$

$$U_0 = \vec{1} = X_0$$

$$U_1 = X_1 - \text{Proj}_{U_0}(X_1) \cdot U_0$$

$$U_2 = X_2 - \text{Proj}_{U_0}(X_2)\, U_0 - \text{Proj}_{U_1}(X_2) \cdot U_1$$

$$\vdots$$

$$U_d = X_d - \text{Proj}_{U_0}(X_d) \cdot U_0 - \text{Proj}_{U_1}(X_d)\, U_1 - \ldots -$$

Original

$$\text{Proj}_{U_{d-1}}(X_d)\, U_{d-1}$$

$$U_i^T U_j = 0$$

orthogonal $\Rightarrow \theta = 90^0 = \pi/2$

original data vector/col

$$\mathrm{Proj}_{U_{d-1}}(X_d)\, U_{d-1}$$

$$\begin{bmatrix} X_0 = U_0 \\ X_1 = U_1 + \mathrm{Proj}_{U_0}(X_1) \cdot U_0 \\ X_2 = U_2 + \mathrm{Proj}_{U_0}(X_2) \cdot U_0 + \mathrm{Proj}_{U_1}(X_2) \cdot U_1 \\ \vdots \\ X_d = U_d + \underbrace{\mathrm{Proj}_{U_0}(X_d)\, U_0 + \mathrm{Proj}_{U_1}(X_d) \cdot U_1}_{\text{matrix equation}} + \cdots + \mathrm{Proj}_{U_{d-1}}(X_d)\, U_{d-1} \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} X_0 & X_1 & \cdots & X_d \end{bmatrix}}_{\tilde{D}} = \underbrace{\begin{bmatrix} U_0 & U_1 & \cdots & U_d \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 1 & P_{10} & P_{20} & \cdots & P_{d0} \\ 0 & 1 & P_{21} & \cdots & P_{d1} \\ 0 & 0 & 1 & P_{31} \cdots & P_{d2} \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}}_{R}$$

Orthogonal columns

all the projection scalars

Upper triangular matrix

$$P_{10} = \mathrm{Proj}_{U_0}(X_1)$$

$$P_{ij} = \mathrm{Proj}_{U_j}(X_i)$$

(Vs. numpy) ⟵

(Vs. numpy) ⟵

→ $Q$ is orthogonal

orthogonal col · unit vector

Ours

$$\tilde{D} = QR$$

numpy

$$Q' = \underbrace{\begin{bmatrix} v_0 & v_1 & \cdots & v_d \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 1/\|v_0\| & & 0 \\ & 1/\|v_1\| & \\ 0 & & \ddots & 1/\|v_d\| \end{bmatrix}}_{\Delta^{-1}}$$

Orthogonal matrix

$$\tilde{D} = Q'R'$$

$$\cancel{R' = \begin{bmatrix} \|v_0\| & p_{10} & p_{20} & \cdots & p_{d0} \\ 0 & \|v_1\| & p_{21} & \cdots \\ & & \ddots & \ddots & \|v_d\| \end{bmatrix}}$$

INCORRECT

fixed after lecture

$$R' = \Delta \cdot R$$

$$R' = \begin{bmatrix} \|v_0\| & & \\ & \|v_1\| & \\ & & \ddots & \|v_d\| \end{bmatrix} R$$

CORRECT

---

You can predict $\hat{Y}$ for all the points.

Q: what about $\vec{\omega}$? and $b$ → Part of $\omega$ as $\omega_0$

Optimization objective:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Col view

$$\|Y - \hat{Y}\|^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Col view

$$\|Y - \hat{Y}\|^2$$

$$J = \|Y - \hat{Y}\|^2 \quad \in \mathbb{R}^n$$

$$= (Y - \hat{Y})^T (Y - \hat{Y})$$

$$= Y^T Y - 2Y^T \hat{Y} + \hat{Y}^T \hat{Y}$$

$$\hat{Y} = w_0 X_0 + w_1 X_1 + \cdots + w_d X_d$$

$$\hat{Y} = \tilde{D}\vec{w}$$

$$\tilde{D} = \begin{bmatrix} X_0 & X_1 & \cdots & X_d \end{bmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix}$$

$$J = Y^T Y - 2Y^T (\tilde{D}\vec{w}) + (\tilde{D}\vec{w})^T (\tilde{D}\vec{w})$$

$$J = \|Y\|^2 - 2\vec{w}^T (\tilde{D}^T Y) + \vec{w}^T (\tilde{D}^T \tilde{D}) \vec{w}$$

$$\frac{\partial J}{\partial \vec{w}} = -2\tilde{D}^T Y + 2(\tilde{D}^T \tilde{D})\vec{w} = 0$$

$$(\tilde{D}\vec{w})^T Y$$

$$\frac{\partial J}{\partial \vec{\omega}} = -2\underline{\tilde{D}}^T Y + 2\left(\tilde{D}^T\tilde{D}\right)\vec{\omega} = 0$$

$$\left(\tilde{D}\vec{\omega}\right)^T Y$$

$$\vec{\omega}^T \tilde{D}^T Y$$

$$\boxed{\left(\tilde{D}^T\tilde{D}\right)\vec{\omega} = \tilde{D}^T Y}$$

Normal equations

$\tilde{D} \in n \times (d+1)$

$\tilde{D}^T \in (d+1) \times n$

$\tilde{D}^T\tilde{D} = (d+1) \times (d+1)$

direct approach:

$$\boxed{\vec{\omega} = \left(\tilde{D}^T\tilde{D}\right)^{-1} \tilde{D} Y}$$

QR - approach to solve for $\vec{\omega}$

$$\boxed{\tilde{D} = QR}$$

$$Q^T Q = \Delta = \begin{bmatrix} \|v_0\|^2 & & & 0 \\ & \|v_1\|^2 & & \\ & & \ddots & \\ 0 & & & \|v_d\|^2 \end{bmatrix}$$

$$\left(\tilde{D}^T\tilde{D}\right)\vec{\omega} = \tilde{D}^T Y$$

$$\Delta^{-1} = \begin{bmatrix} 1/\|v_0\|^2 & & \\ & 1/\|v_1\|^2 & \\ & & 1/\|v_d\|^2 \end{bmatrix}$$

$$(QR)^T(QR)\vec{\omega} = (QR)^T Y$$

$$R^T Q^T Q R \vec{\omega} = (QR)^T Y$$

$$\cancel{R^T}\Delta R\vec{\omega} = \cancel{R^T} Q^T Y$$

$\Delta R \vec{\omega} = Q^T Y$

$R\vec{\omega} = \underline{\Delta^{-1} Q^T Y}$

Upper triangular matrix

Solve for $\vec{\omega}$ using backsolve!

$$\underbrace{\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}}_{R} \underbrace{\begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}}_{\vec{\omega}} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

last row $\quad w_2 = d_3$ !

2nd last row : $\quad w_1 + c \cdot w_2 = d_2$

$w_1 + c d_3 = d_2$

$w_1 = d_2 - c d_3$

1st row ! $\quad w_0 + a w_1 + b w_2 = d_1$

$w_0 = d_1 - a w_1 - b w_2$

$w_0 = d_1 - a(d_2 - c d_3) - b d_2$

$$\begin{bmatrix} 1 & 0.1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Q