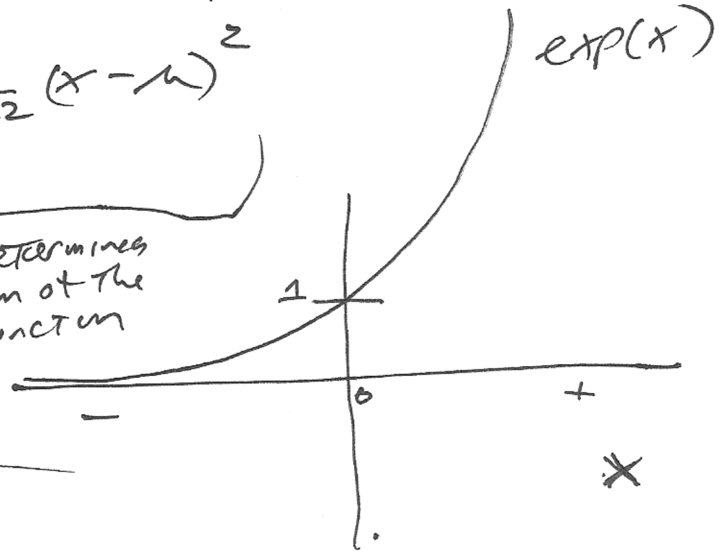
Univariate Gaussian / Normal distribution

$$N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Just a normalization constant so we integrate to 1

This determines form of the function



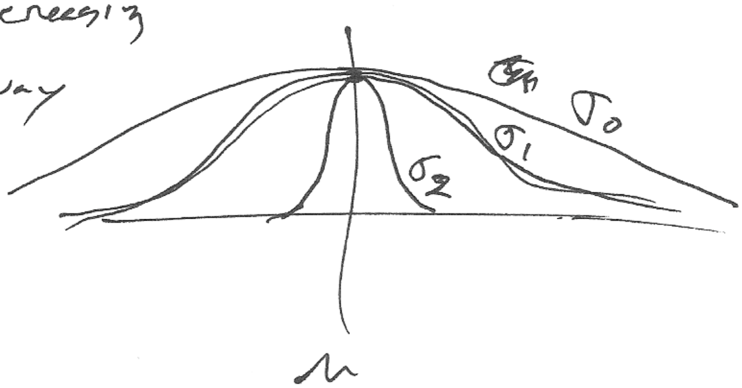
$(x-\mu)^2$ measure squared distance ~~from~~ of x from μ

- always positive
~~for $N(\mu,\sigma)$ distribution bounded above~~
~~for negative #~~
- for non positive #s which $\frac{-1}{2\sigma^2}(x-\mu)^2$ is, bounded by 1, at $x=\mu$

as distance increases, $-(x-\mu)^2$ ~~increases~~ decreases & density goes down.

so max at $x=\mu$, decreasing as you get farther way

σ^2 determines how fast the drop off happens



Holding $\frac{(x-\mu)^2}{2}$ fixed, as σ increases, then $-\left(\frac{x-\mu^2}{2\sigma^2}\right)$ is smaller magnitude, (closer to zero) so drop off is slower.

$\sigma_0 > \sigma_1 > \sigma_2$

μ tells location of highest density

σ tell spread of density (how tightly/loosely it is clustered about μ)

multivariate Gaussian / normal distribution

generalization of univariate case to k dimensions ~~[illegible]~~

~~[illegible]~~

distribution still describes a point location of highest concentration and spread of data about that location, but the representation for spread of the data is more complicated since you have to describe how it spreads in different directions (in ~~[illegible]~~ $\mathbb{R}^k$)

$$N(x; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \underbrace{e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}_{\text{again, look at this part}}$$

show how specializes to univariate case $k=1$ note $\Sigma = \sigma^2$.

Case 1

let $\Sigma = \sigma^2 I$

Then we have $-\frac{1}{2\sigma^2}\underbrace{(x-\mu)^T(x-\mu)}_{\text{squared distance from } \mu}$

very similar to univariate case; we have max density when $x=\mu$, and drop off of density away from that depending on value of σ^2. In this case, drop off is same (radially symmetric) in all directions of $\mathbb{R}^k$

Case 2

now consider

$$\Sigma = \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix}$$

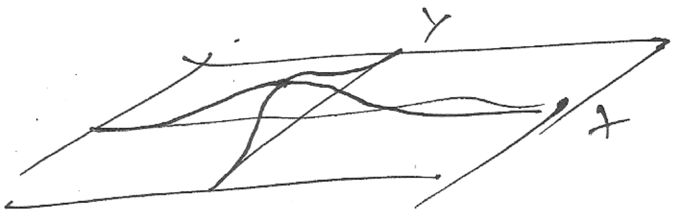
note: bivariate ($k=2$) [so we can draw a picture of it]

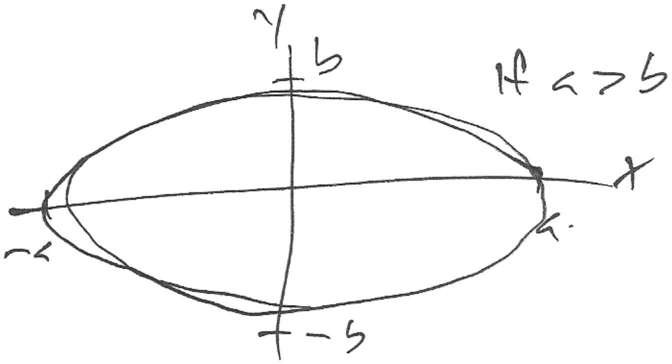
Let's also set $\mu = 0$ to make life easier ...

$$e^{-\frac{1}{2}\begin{pmatrix}x & y\end{pmatrix}\begin{pmatrix}1/a^2 & 0\\ 0 & 1/b^2\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}} \Rightarrow e^{\frac{-1}{2a^2}x^2}\, e^{\frac{-1}{2b^2}y^2}$$

Let $x=0$, reduces to $e^{\frac{-1}{2b^2}y^2}$ univariate normal in y direction

Let $y=0$, reduces to $e^{\frac{-1}{2a^2}x^2}$ univariate normal in x





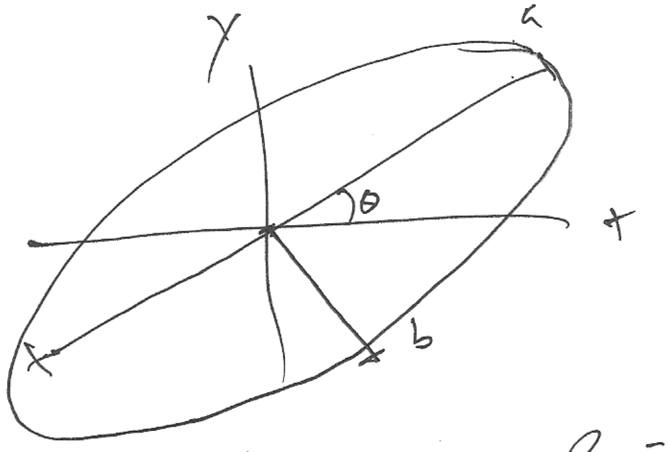
if $a > b$

case 3

"general" 2x2 Σ general = ~~positive~~ symmetric, ~~positive definite~~ $\det\Sigma > 0$

eigendecomposition, can decompose into

$$\Sigma = R\begin{bmatrix}a^2 & 0\\ 0 & b^2\end{bmatrix}R^T$$ R is rotation matrix



ellipse is rotated, depending on $R = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$

In general for higher k

Σ k×k symmetric, pos definite covariance matrix

do eigendecomposition

$$\Sigma = U D U^T$$

$$D = \begin{bmatrix} d_1^2 & & 0 \\ & d_2^2 & \\ 0 & & \ddots \; d_k^2 \end{bmatrix}$$

$$U = [\hat{u}_1 \; \hat{u}_2 \; \cdots \; \hat{u}_k]$$ orthogonal unit vectors

d_i describes spread of distribution/mass/data in direction $\hat{u}_i$

constant prob contours are ellipsoids

if $d_1 > d_2 > \ldots d_k$

Then major axis of ellipsoid is $\hat{u}_1$

next most is $\hat{u}_2$, and so on . . .