Gaussian Random Variables and The Kalman Filter

Ron Parr
CPS 1/296

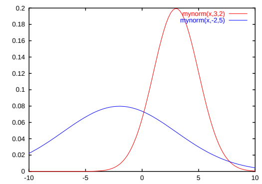
With some content adapted from Andrew Ng, Lise Getoor, and Tom Dietterich

Multivariate normal distribution

- also called multivariate gaussian
- First, recall the univariate normal distribution:

$$p(x;\mu,\sigma) = \frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$$

- where μ is the mean and σ^2 is the variance



Multivariate normal distribution

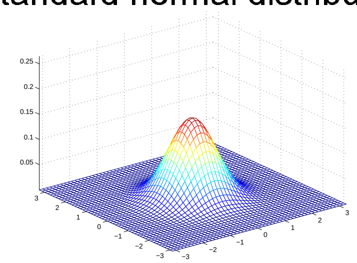
- A 2-dimensional gaussian is defined by a mean vector $\mu = (\mu_1,\mu_2)$ and a covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma^2_{1,1} & \sigma^2_{1,2} \\ \sigma^2_{2,1} & \sigma^2_{2,2} \end{bmatrix}$$

- where $\sigma^2_{i,j} = E[(x_i-\mu_i)(x_j-\mu_j)]$
  - is the variance if $x_i = x_j$
  - covariance if $x_i \neq x_j$
- where
  - is the variance if $x_i = x_j$
  - covariance if $x_i \neq x_j$

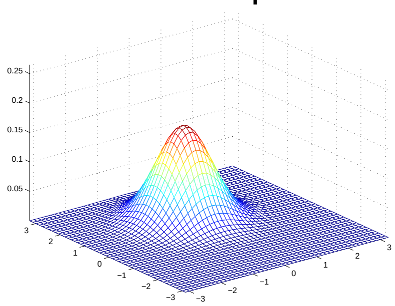
$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)\,|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

Standard normal distribution



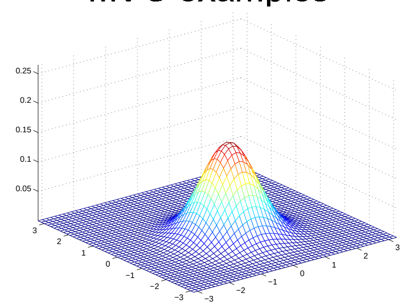
- We get the standard normal for Σ = the identity matrix $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mu = (0,0)$

MVG examples



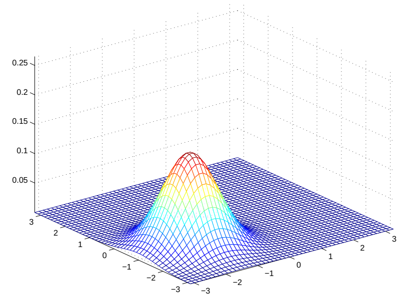
$\mu = (1, 0)$ $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

MVG examples



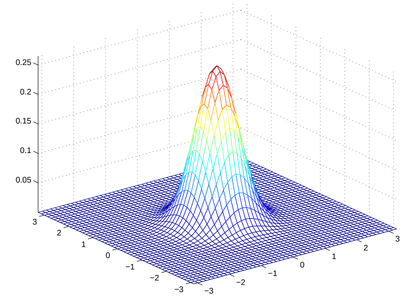
$\mu = (-0.5, 0)$ $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

MVG examples



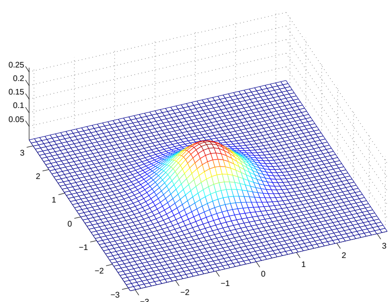
$\mu = (-1,-1.5) \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

MVG examples



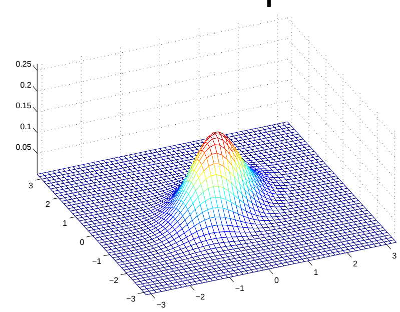
$\mu = (0,0) \quad \Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}$

MVG examples



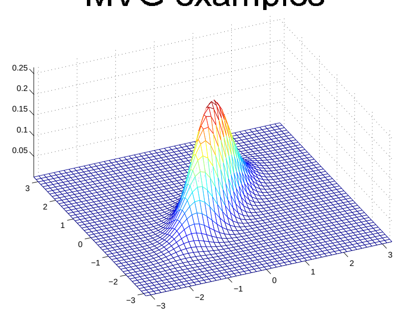
$\mu = (0,0) \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

MVG examples



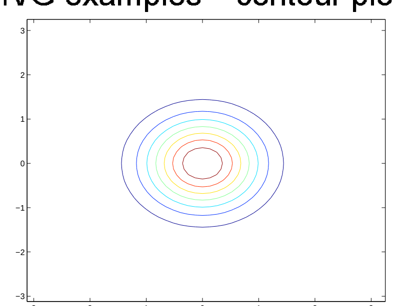
$\mu = (0,0) \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

MVG examples



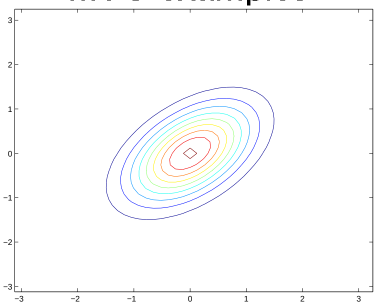
$\mu = (0,0) \quad \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$

MVG examples – contour plots



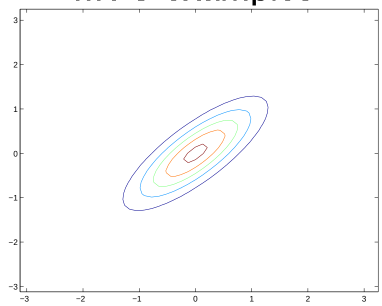
$\mu = (0,0) \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

MVG examples



$\mu = (0,0) \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

MVG examples



$\mu = (0,0) \quad \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$

Multivariate normal distribution

- We can generalize this to n dimensions
- parameters
  - mean vector $\mu \in \Re^n$
  - a covariance matrix $\Sigma \in \Re^{n \times n}$, where $\Sigma \geq 0$ is symmetric and positive semi-definite
- Written $N(\mu, \Sigma)$, density is

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

- where $|\Sigma|$ is the determinant of the matrix Σ
- For $X \sim N(\mu, \Sigma)$
  - $E[X] = \int_x x\, p(x;\mu,\Sigma)\, dx = \mu$
  - $Cov(X) = E[XX^T] - (E[X])(E[X])^T = \Sigma$

Summary of Useful Gaussian Properties

- Linear combinations of Gaussians are Gaussian
- Conditionals of Gaussians are Gaussian
- Marginals of Gaussians are Gaussian
- Relatively compact: $O(n^2)$ space

Linear Gaussian State Transition Models

$$x_t = A_t x_{t-1} + B_t u_t + \varepsilon_t$$

- x_t = column vector of continuous state variables
- A_t = Linear state dynamics matrix
- U_t = column vector of control inputs
- B_t = Linear control model
- ε_t = Gaussian noise (mean 0 jointly Gauusian random vector with cov= R_t)
- R,A,B often stationary

(Optional: U_t, B_t)

Interpreting the Dynamics

$$x_t = A_t x_{t-1} + B_t u_t + \varepsilon_t$$

- Given any particular value of x_{t-1}, x_t is Gaussian random vector
- If x_{t-1} is a Gaussian random vector, then x_t is also a Gaussian random vector (linear combination of Gaussians)

Computing The Distribution

- For an HMM:

$$P(x_t) = \sum_{x_{t-1}} P(x_t \mid x_{t-1}) P(x_{t-1})$$

- Generalizing:

$$p(x_t) = \int_{x_{t-1}} p(x_t \mid x_{t-1}) p(x_{t-1}) dx_{t-1}$$

For Gaussians

$$p(x_t) = \int_{x_{t-1}} p(x_t \mid x_{t-1}) p(x_{t-1}) dx_{t-1}$$

Jointly Gaussian

Jointly Gaussian

Marginal of jointly Gaussian RVs

Algorithmic view of means

- Assume x_{t-1} is mean vector at time t-1
- x_t^- is prediction at time t

$$x_t^- = A x_{t-1}$$

Algorithmic view of covariances

- Assume Σ_{t-1} is cov matrix at time t-1
- Σ_t^- is prediction covariance at time t

$$\Sigma_t^- = A \Sigma_{t-1} A^T + R$$

Observation Model

$$z_t = C_t x_t + \delta_t$$

- C_t is a linear observation model
- δ_t is Gaussian random vector w/cov Q_t

- C,Q often stationary

Conditioning on observations

$$x_t = x_t^- + K(z_t - C x_t^-)$$
$$\Sigma_t = (I - K_c C)\Sigma_t^-$$
$$K = \Sigma_t^- C^T (C \Sigma_t^- C^T + Q)^{-1}$$

- Called the “correction step”
- Result of (omitted, tedious) derivation from conditioning on observation variable
- Notice “missing” dependence in covariance

KF Derivation Notes

- Kalman filter derivation is a tedious, but straightforward calculus exercise
- Can be derived in multiple ways
  - Marginalization, conditioning
  - Show KF is estimate with minimum error

KF Notes

- Advantages of Kalman filter:
  - Simple matrix-vector implementation
  - State representation remains constant size
- Disadvantages of Kalman filter:
  - Makes strong assumptions
  - Requires matrix inversion

Questions

- What if dynamics aren't linear?
- What if distributions aren't Gaussian?
- What if there are discrete variables?
- Partly addressed by:
  - Particle filter
  - Rao Blackwellized particle filter
  - Extended Kalman filter
  - Unscented filter