C20220 Notes 8: Revision ideas

Daniel Richardson, Norwood House 6.29, email masdr@bath.ac.uk

April 30, 2009

1 Definitions

1. Vector space over $\mathbf{R}$
2. $C[a, b]$. Definition of addition and scalar multiplication for $C[a, b]$.
3. Dot product in $\mathbf{R}^N$. $\mid X \mid$. $d(X, Y)$. $X \perp Y$. $Proj_X Y$, forX, Y in $\mathbf{R}^N$.
4. Inner product. Inner product space. Definition of $\mid X \mid$, etc for X in inner product space. (as previous item.)
5. $< f, g >$, for $f, g \in C[a, b]$. Hence $\mid f \mid$, etc.
6. Order k Fourier approximation to $f \in C[-\pi, \pi]$.
7. Discrete Fourier transform in one and two dimensions. Also inverse discrete Fourier transform.,
8. filter, linear filter.
9. $f * g$, convolution of f with g, in one and two dimensions.
10. Global/local min/max of $f : \mathbf{R}^D \to \mathbf{R}$.
11. ∇f, level surface
12. Neural net
13. Random variable. Universe, event.
14. Sample
15. Conditional probability $P(X \in A \mid X \in B)$.
16. expected value of $f(X)$, where f maps the universe of X into the reals.
17. Mean, variance, standard deviation, covariance.

18. Probability density function $d(x)$ for random variable X with universe $\mathbf{R}$ or $\mathbf{R}^D$.

19. Normal distribution.

20. Gaussian distribution, mixture of Gaussians.

21. independence of random variables X and Y.

2 Theorems and Formulae

1. Theorem 1. If $X_1, \ldots, X_k, Y$ are in vector space V with inner product $<,>$, and $X_1, \ldots, X_k$ are mutually orthogonal and non zero, then the best approximation to Y in $SPAN(X_1, \ldots, X_k)$ is $c_1X_1 + \ldots c_kX_K$, where, for $i = 1, ...k$,

   $c_i =< X_i, Y > / < X_i, X_i >$.

2. (Fourier approximation theorem).

   $F_k(t) = a_0/2 + \Sigma_{n=1}^k a_n cos(nt) + b_n sin(nt)$

   where $a_n = (1/\pi) \int_{-\pi}^{\pi} f(t)cos(nt)dt$, and

   $b_n = (1/\pi) \int_{-\pi}^{\pi} f(t)sin(nt)dt$, is the best approximation to $f(t)$ on $[-\pi, \pi]$ in

   $SPAN(1, cos(t), sin(t), ..., cos(nt), sin(nt))$.

3. $f * g$ the convolution of f with g is linear in both f and g.

4. (Convolution Theorem) If F is the discrete Fourier transform of f, and G is the discrete Fourier transform of g, then FG is the discrete Fourier transform of $f * g$.

5. (Deconvolution Theorem) See notes.

6. (Gradient Theorem). If $f : \mathbf{R}^D \to \mathbf{R}$ is differentiable, and α is a local minimum or maximum, then $\nabla f(\alpha) = 0$.

7. Bayes Theorem. See notes.

8. Central Limit Theorem. See notes.

3 Ideas and Techniques

I have not included everything, but this should give you some framework.

4 Approximation

1. Polynomial curve fitting. Given data $= (x_1, y_1), \ldots, (x_n, y_n)$, find a polynomial $p(x)$ of degree d to minimise

   $\Sigma_{i=1}^n (p(x_i) - y_i)^2$

2. Vector approximation

3. Order k Fourier approximation to a function $f : [-\pi, \pi] \to \mathbf{R}$.

Notes:

1. Fourier approximation is stable under small perturbations of the function; however polynomial curve fitting is not stable under small perturbations of the data.That is, for polynomial curve fitting, a small change in the data can result in a huge change in the coefficients of the polynomial.

2. For polynomial curve fitting, if the degree d is at least $n-1$, and the x values are distinct, there exists a unique polynomial which goes exactly through every point. However, if there is some randomness in the data, it may be better to choose a much smaller degree. That is, sometiems simple models are better than complicated ones.

5 Optimisation and minimisation

Suppose $f : \mathbf{R}^D \to \mathbf{R}$.

1. Direct method, by solving $\nabla f = 0$. This is D equations in D unknowns, but is not linear unless f is quadratic. But this does polynomial curve fitting and vector approximation problems.

2. Steepest gradient descent.

3. Otimisation of neural nets.

6 Image transformation using convolution and masks

7 Statistics

1. Given sample of random variable X in $\mathbf{R}^D$, compute Gaussian model. Or mixture of Gaussians if given more than one cluster.

2. Given sample of continuous random variable, construct probability density function.

3. Bayesian computation. Example: given $P(Sy \mid PF), P(Sy \mid \textit{ no } PF), P(PF)$, compute $P(PF \mid Sy)$.