Lecture 1: Intro + Some Review

Welcome to AMATH 482/582

* Make sure you read the syllabus doc on Canvas. It has crucial info about the course

Course Instructor:

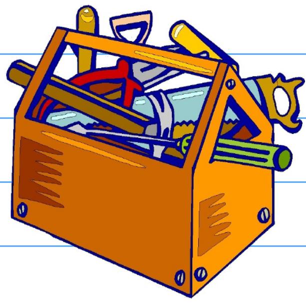
Bamdad Hosseini
Office: Lewis Hall 220
Email: bamdadh@uw.ed

Teaching Assistants:

Katherine Owens
Email: lacyk3@uw.edu

Roman Levin
Email: rilevin@uw.edu

* Think of this course as a toolbox for analyzing data.

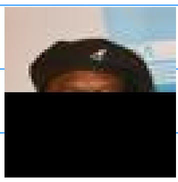


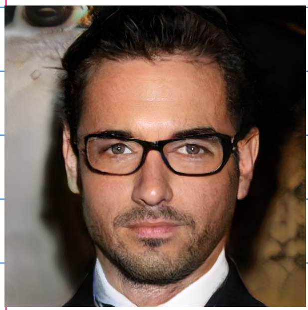
Rise of DS & ML:

* In October 2015, in a match against Fan Hui, the original AlphaGo became the first computer Go program to beat a human professional Go player without handicap on a full-sized 19×19 board.[5][6] In March 2016, it beat Lee Sedol in a five-game match, the first time a computer Go program has beaten a 9-dan professional without handicap.[7] Although it lost to Lee Sedol in the fourth game, Lee resigned in the final game, giving a final score of 4 games to 1 in favour of AlphaGo. In recognition of the victory, AlphaGo was awarded an honorary 9-dan by the Korea Baduk Association.[8] The lead up and the challenge match with Lee Sedol were documented in a documentary film also titled *AlphaGo*,[9] directed by Greg Kohs. The win by AlphaGo was chosen by *Science* as one of the Breakthrough of the Year runners-up on 22 December 2016.[10]

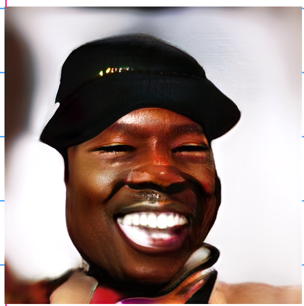
AlphaGo

*


*   

Impressive results!

Beware the failures!

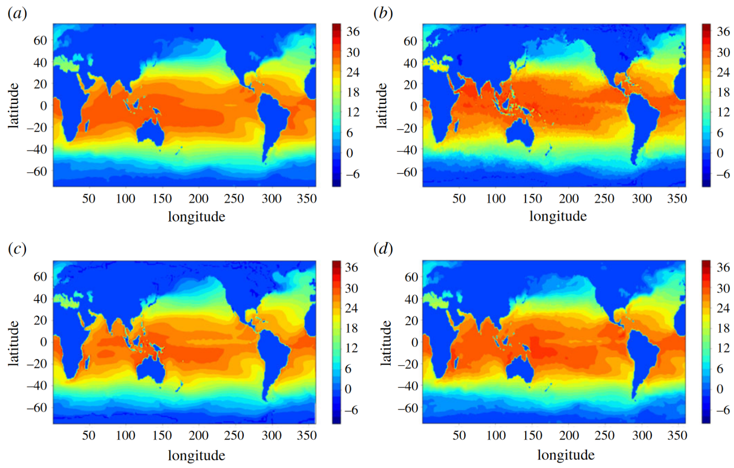
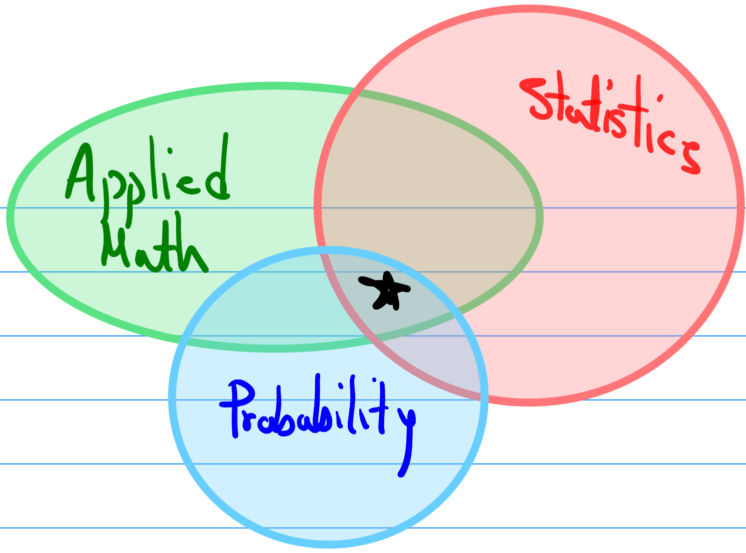
* 

Figure 5. Contour plots for various forecasts ((*a*) POD-RKHS, (*b*) HYCOM, (*c*) CESM, (*d*) truth), indicating how low-dimensional manifold-based emulation reduces the ability to capture fine-scaled features in the flow-field. HYCOM, with the finest resolution, is seen to capture small-scale information most accurately. Note, however, that the POD-based emulation framework is competitive in an averaged sense, as seen through RMSE metrics. (Online version in colour.)

* DS is widely used in natural sciences.

- These impressive results are built upon the same ideas & tools that you will see in this course.
- By the end of the quarter you will be knowledgeable to the point of understanding some cutting edge research.



Review of Basic Concepts from Stats & Probability

Let us consider a discrete random variable X with possible states $x_1, \dots, x_n$, with corresponding probabilities $p_1, \dots, p_n$. That is,

$$\mathbb{P}(X = x_j) = p_j.$$

At an abstract level we think of X as a function

$$X : \Omega \longrightarrow \{x_1, \dots, x_n\}$$

where Ω is the "set of possible outcomes".

ex: Let X be the outcome of a dice roll.

Then $X : \Omega \longrightarrow \{1, 2, \dots, 6\}$ & $p_1 = p_2 = \dots = p_6 = 1/6$.

Given a discrete r.v. X we define it's expected value (or mean) as

$$\mathbb{E}(X) := \sum_{j=1}^{n} p_j x_j$$

along with it's variance

$$\mathbb{V}\mathrm{ar}(X) := \mathbb{E}((X - \mathbb{E}(X))^2).$$

Throught this course we often need r.v. that are real or vector valued. In this case we take the abstract def^n $X : \Omega \to \mathbb{R}^d$ for $d \geq 1$.

We say $\pi : \mathbb{R}^d \to \mathbb{R}$ is a probability density function (PDF) for the r.v. X if

$$\mathbb{P}(X \in A) = \int_A \pi(x)\, dx,$$ d-dim integral.

for any open set $A \subseteq \mathbb{R}^d$.

We can now define the expected value (expectation) of X as

$$\mathbb{R}^d \ni \mathbb{E}(X) := \int_{\mathbb{R}^d} x\, \pi(x)\, dx$$

Along with the Covariance operator (matrix)

$\mathbb{R}^{d\times d} \ni \mathrm{Cov}(X) := \int_{\mathbb{R}^d} (x-\mathbb{E}(X))(x-\mathbb{E}(X))^T \pi(x)\, dx$

Recall that $x \in \mathbb{R}^d$ (column vec) & so $xx^T \in \mathbb{R}^{d\times d}$.
where x^T denotes the transpose of x.

When X is real valued, $X: \Omega \to \mathbb{R}$ then

$$\mathrm{Cov}(X) = \int_{\mathbb{R}^d} (x-\mathbb{E}(X))^2 \pi(x)\, dx$$

$$\equiv \mathbb{V}\mathrm{ar}(X).$$

Yields the definition of Var for real valued r.v.

If you don't like the integrals consider the following empirical approximations:

suppose X is an $\mathbb{R}^d$ valued r.v & let

$x_1, \ldots, x_N$ denote N independent & identical realizations of it (ie. X is generated N times). Then the empirical average

$$S_N(X) := \frac{1}{N}\sum_{j=1}^{N} x_j$$

approximates the expectation (look up law of large numbers)

$$S_N(X) \longrightarrow \mathbb{E}(X) \quad \text{as } N \to \infty$$

which in turn implies

$$\frac{1}{N}\sum_{j=1}^{N}(x_j - S_N(x))(x_j - S_N(x))^T \longrightarrow \mathrm{Cov}(X) \quad \text{as } N \to \infty.$$

We used the term "independent" above. This has a technical def^n that we don't want to get into. But

loosely speaking two r.v.

$$X : \Omega \to \mathbb{R}^d, \quad Y : \Omega \to \mathbb{R}^d$$

are independent if their outcomes are disjoint events (independent), i.e,

Given sets $A, B \subset \mathbb{R}^d$ we have

$$\mathbb{P}(X \in A \text{ and } Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B)$$

<u>ex</u>: Consider two simultaneous coin flips.
Then $\mathbb{P}(\text{Coin}_1 = \text{head and Coin}_2 = \text{head})$

$$= \mathbb{P}(\text{coin}_1 = \text{head})\,\mathbb{P}(\text{coin}_2 = \text{head}) = 1/4$$

similarly for other outcomes.

If X, Y are independent $\mathbb{R}^d$ valued r.v.
then $\mathbb{E}(X^TY) = \mathbb{E}(X)^T\mathbb{E}(Y)$.

Let us further extend the notion of the Covariance to multiple r.v. & write

exp. wrt both X, Y

$$\mathrm{Cov}(X,Y) := \mathbb{E}(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))^T$$

$$= \mathbb{E}\left[XY^T - \mathbb{E}(X)Y^T - X\mathbb{E}(Y)^T + \mathbb{E}(X)\mathbb{E}(Y)^T\right]$$

(linearity)

$$= \mathbb{E}[XY^T] - 2\,\mathbb{E}(X)\mathbb{E}(Y)^T + \mathbb{E}(X)\mathbb{E}(Y)^T$$

$$= \mathbb{E}[XY^T] - \mathbb{E}(X)\mathbb{E}(Y)^T$$

Observe that the expectation is "linear" that is, given $a \in \mathbb{R}^d$ & $b \in \mathbb{R}$ we have

$$\mathbb{E}(a^TX + b) = a^T\mathbb{E}(X) + b.$$

To conclude, if X, Y are independent $\mathbb{R}^d$ valued r.v's then

$$\mathrm{Cov}(X,Y) = 0.$$

zero matrix!

We say that a sequence of r.v's $X_1, \dots, X_K$ are independent & identically distributed (i.i.d.) if the pairs $(X_i, X_j)_{i,j=1}^{K}$ are independent but the $\{X_j\}_{j=1}^{K}$ have the same PDEs

Then we obtain the following:

Suppose X is an $\mathbb{R}^d$ valued r.v. so that its components (coordinates) are i.i.d.

Then, $Cov(X) = Id$. Identity matrix.