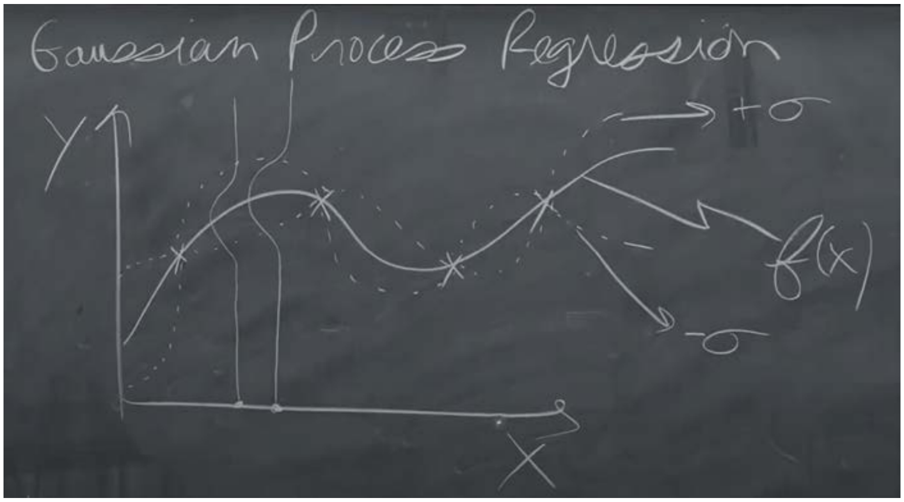
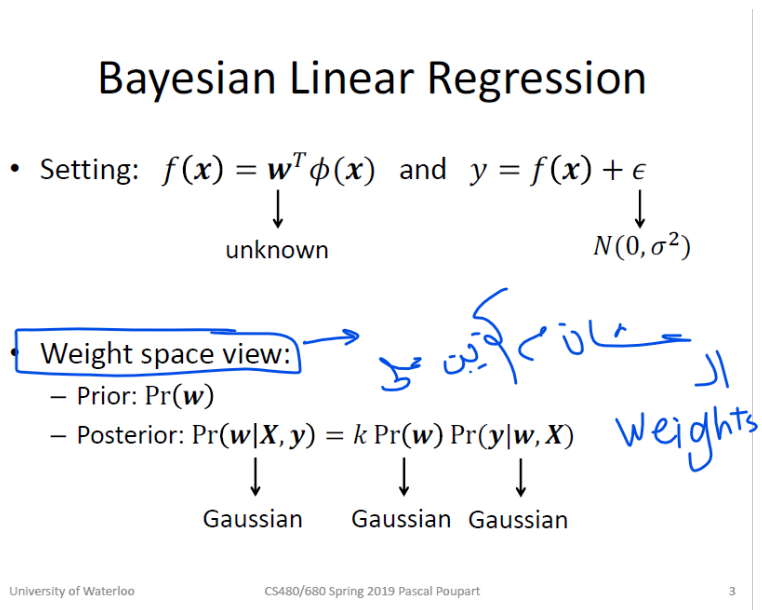
انهارده هنتكلم علي gaussian process وده موديل يعني .. و ده corresponds to baysian linear regression when we use kernels فدلوقت هنشوف ميزة الكيرنال ... فهي بتخلينا نقدر نحسب الدوت برضودكت من غير ما نروح لل new high dimensional space فنقدر نعمل ده بسرعه .. فتعال نشغل اكلام ده علي generalized linear model واول واحد هنتكلم عليه هو ال baysian linear regression ومنين ما بتستخدم الكيرنالز في الكونتكست ده .. ده بيودينا ناحية ال Gaussian process . ده اول جزء ... جاوسين بروسيسز برضو بتبقا equivilant to a neural network اللي ليها one hidden layer بس بتبقا infinitely large ... ففي ربط ما بين الكلاسيك تكنيكس زي مثلا ال baysian linear regression اللي هنسميها دلوقت جاوسين بروسيس .. فده بيبقا مكافئ ل Bayesian learning in a neural network with 1 hidden layer whenever you've got an infinitly large hidden layer

طيب عشان نوضح ايه هي الجاوسين بروسيس خلينا نرسم صوره و بعدين نتكلم .....

احنا مهتمين بالرجريشن وعندنا داتا .. الانبوت هو الإكس اكسيس و الاوتبوت هو الواي أكسيس .. الفانكشن اللي انت عاوز تكتشفها هي الفانكشن اللي الدكتور راسمها .. انت عندك شوية داتا علي الخط اللي انت نفسك تجيبه ... احنا هنفترض ان الداتا مفهاش نويز و الواي الي بتقيسها ملهاش نويز .. فدلوقت لو بتعمل لينير ريجريشن مش هتقدر تجيب الفانكشن ديه عشان ديه فانكشن نان لينير ..



طيب لما كنا بنتكلم علي البايزين لينير ريجريشن .. لما كنا بنشتغل فيه فكنا بنفترض ان الفانكشن f(x) بتبقا linear combinations of phi(x) و الأوتبوت f(x) ده كان معاه شوية noise فطلعلك ال y ... بص عالسلايد



دلوقت هنبص علي function space view فبدل ما هنفكر في ال w هي الآن نوون .. دلوقت هنا الحاجه الي احنا بنحاول نعملها estimation هي الفانكشن نفسها f(x) .. الفانكشن نفسها بتعتمد علي شوية بارمترز w عادي ... بس هنا هنتعامل مع ال f علي اساس انها ال unknown .. فلو هنعمل baysian linear regression هنبدأ بال prior over a function ده بيبقا زي اللي ف السلايد كدا ...

Bayesian Linear Regression

- Setting: $\underbrace{f(\boldsymbol{x})}_{\text{unknown}} = \boldsymbol{w}^T\phi(\boldsymbol{x})$ and $y = f(\boldsymbol{x}) + \underbrace{\epsilon}_{N(0,\sigma^2)}$
- Function space view:
  - Prior: $\underbrace{\Pr\big(f(\boldsymbol{x}_*)\big)}_{\text{Gaussian}} = \int_{\boldsymbol{w}} \underbrace{\Pr(f|\boldsymbol{w},\boldsymbol{x}_*)}_{\text{Deterministic}} \underbrace{\Pr(\boldsymbol{w})}_{\text{Gaussian}}\, d\boldsymbol{w}$
  - Posterior: $\underbrace{\Pr(f(\boldsymbol{x}_*)|\boldsymbol{X},\boldsymbol{y})}_{\text{Gaussian}} = \int_{\boldsymbol{w}} \underbrace{\Pr(f|\boldsymbol{w},\boldsymbol{x}_*)}_{\text{Deterministic}} \underbrace{\Pr(\boldsymbol{w}|\boldsymbol{X},\boldsymbol{y})}_{\text{Gaussian}}\, d\boldsymbol{w}$

newname
10/11/2021 7:59:21 AM

ديترمنستيك عشان انت لو
عرفت ال w هتبقا عارف ال f
فبالتالي انت معندكش
uncertainty ..

Gaussian Process

- According to the function view, there is a Gaussian at $f(\boldsymbol{x}_*)$ for every $\boldsymbol{x}_*$. Those Gaussians are correlated through w.

- What is the general form of $\Pr(f)$ (i.e., distribution over functions)?
- Answer: **Gaussian Process** (infinite dimensional Gaussian distribution)

الدكتور بيقول ان في correlation between gaussians through w .. ومن هنا بيجي ال kernel .. عشان لو انت لاحظت نقطه وانت عارف ان الفانكشن لازم تعدي بيها .. وكمان انت عارف ان في correlation between gaussians بالنسبه لل weights فانت لو خدت نقطه قيربه للنقطه للي انت بتلاحظها دلوقت .. هتلاحظ uncertainty صغير جداً ... وده بيروح لفكرة ان لو انت عندك condition on a value for a given point يبقا لو عندك نقطه فعلاً قريبه .. هيبقا الاوتبوت بتاعها قريب جداً فبرضو هيبقا عندك little uncertainty فعشان كدا في الرسمه اللي الدكتور راسمها .. هتلاقي ان الكيرفين بتوع ال +sigma and –sigma بيبعدو عن بعض .. بس في الاول هم بيبقو قريبين من بعض وبعدين بيبدأو بالتدريج يبعدوا ... فانت فعلياً هتتوقع little uncertainty في الاول يعني ... الجاوسين بروسيس هي infite number of Gaussian distributions all correlated together .... فده معناه انك عندك infinite multivariate Gaussian .. وده اللي اسمو Gaussian process ....

Gaussian Process

newname
10/11/2021 8:08:52 AM

ازاي نربرزنت الجاوسين ديستربيوشن ..
هنحطلو مين و كوفارينس ...

- Distribution over functions:

$$f(\boldsymbol{x}) \sim GP\big(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big) \; \forall \boldsymbol{x}, \boldsymbol{x}'$$

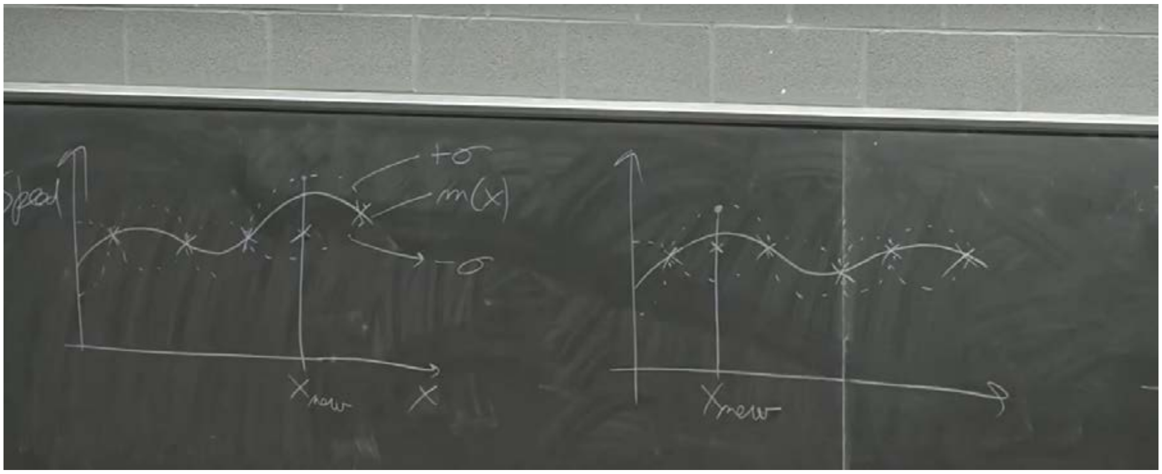
- Where $m(\boldsymbol{x}) = E(f(\boldsymbol{x}))$ is the mean and $k(\boldsymbol{x}, \boldsymbol{x}') = E((f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))$ is the kernel covariance function

Gaussian Process Regression

newname
10/11/2021 8:21:11 AM

الدوت معناها ان الإكس ممكن تبقا اي
حاجه .. لو شلتها وحطيط مكانها إكس يبقا
انت بتتكلم علي إكس واحده بس

- Prior: $\Pr(f(\cdot)) = N(m(\cdot), k(\cdot,\cdot))$
- Likelihood: $\Pr(\boldsymbol{y}|\boldsymbol{X}, f) = N(f(\boldsymbol{X}), \sigma^2 \boldsymbol{I})$
- Posterior: $\Pr(f(\cdot)|\boldsymbol{X}, \boldsymbol{y}) = N\big(\bar{f}(\cdot), k'(\cdot,\cdot)\big)$
  where $\bar{f}(\cdot) = k(\cdot, \boldsymbol{X})(\boldsymbol{K} + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{y}$ and
  $$k'(\cdot,\cdot) = k(\cdot,\cdot) + \sigma^2 \boldsymbol{I} - k(\cdot, \boldsymbol{X})(\boldsymbol{K} + \sigma^2 \boldsymbol{I})^{-1} k(\boldsymbol{X}, \cdot)$$
- Prediction: $\Pr(y_*|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}) = N\left(\bar{f}(\boldsymbol{x}_*), k'(\boldsymbol{x}_*, \boldsymbol{x}_*)\right)$
- Complexity: inversion of $\boldsymbol{K} + \sigma^2 \boldsymbol{I}$ is cubic in # of training points


Speed
$+\sigma$
$m(x)$
$-\sigma$
X_{new}
X
X_{new}