Gaussian Processes and Bayesian Optimization

Quiz, 8 questions

Congratulations! You passed!

Next Item

1 / 1 point

1.
Select properties of nonparametric methods.

- [ ] Prediction model is highly constrained to the specified form.

Un-selected is correct

- [x] Usually learning process is just remembering the dataset.

Correct
Usually non-parametric methods need to remember whole dataset for prediction, while parametric methods need to remember only a fixed set of fitted parameters.

- [ ] Very fast prediction.

Un-selected is correct

- [x] Have a lot of parameters depending of dataset size.

Correct
Non-parametric methods do not assume that there is fixed set of parameters to tune on the training dataset.

---

1 / 1 point

2.
Select stationary Gaussian Processes.

- [ ] $m(x) = x$ and $K(x, x') = min(x, x')$.

Un-selected is correct

- [x] $m(x) = const$ and $K(x, x') = I[x = x']$.

Correct
Mean vector and covariance matrix are the same for every set of points (x_1, ..., x_n).

- [x]

$m(x) = 0$ and $K(x, x') = 1/(1 + (x - x')^2)$.

Correct

$m(x) = const$, $K(x, x') = \tilde{K}(x - x')$.

- [ ] $m(x) = const$

Un-selected is correct

---

1 / 1 point

3.

Choose the time complexity of prediction with Gaussian Process (d — dimension of features, n — number of training objects).

- ( ) $O(nd^2)$.
- ( ) $O(n^2 d)$.
- (•) $O(n^3 + dn^2)$.

Correct

$O(dn^2)$ for computing the matrix C and $O(n^3)$ for inverting.

- ( ) $O(n^3)$.

---

1 / 1 point

4.

Consider the regression problem, $\{x_i, y_i\}_{i=1}^n$ — the training set. Assume that $y_i = f(x_i)$ — realization of stationary Gaussian Process ($m(x)$ — mean function and $m(x) = 0$, $K(x, x') = \tilde{K}(x - x')$ - continuous covariance function with the property $\tilde{K}(t) \to 0$ if $t \to \infty$), x — new object. Select all correct statements about $p(f(x)|f(x_1), f(x_2), ..., f(x_n))$.

- [x] If the object is far from the training sample then the variance of $p(f(x)|f(x_1), ..., f(x_n))$ is close to $K(0)$.

Correct

the formula $\sigma^2 = K(0) - k^T C^{-1} k$ implies this result. If you put x away from all the objects in the training set you get $k \to 0$.

- [ ] Expected value of $p(f(x)|f(x_1), ..., f(x_n))$ is close to 0, if x is close to training sample.

Un-selected is correct

- [x] $p(f(x)|f(x_1), f(x_2), ..., f(x_n))$ is normal distribution.

Correct

All the computations that lead to that result are described in the lecture 3.

- [x] Variance of $p(f(x)|f(x_1), ..., f(x_n))$ is close to 0, if x is close to training sample.

Correct
Continuity of the covariance function implies this result.

- [ ] Expected value of $p(f(x)|f(x_1), ..., f(x_n))$ doesn't depend on $c_{i,j} = K(x_i, x_j)$.

Un-selected is correct

---

1 / 1 point

5.

Consider the following kernel: $\tilde{K}(x_i - x_j) = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right) + s^2\mathbb{I}[x_i = x_j]$. Which statements about this kernel are true?

- [x] Parameters of the model s, l and σ can be optimized with gradient descent.

Correct

- [x] If the object is far from the training sample then the variance of $p(f(x)|f(x_1), ..., f(x_n))$ is close to $K(0)$.

Correct
Because $\tilde{K}(t) \to 0$ if $t \to \infty$

- [ ] $Var(f(x)|f(x_1), ..., f(x_n)) = 0$, if $x = x_i$.

Un-selected is correct

- [x] Higher values of s^2 give you the model that is more robust to the noise in the data.

Correct
This term assumes that there is an unpredictable noise in the data.

---

1 / 1 point

6.

Which of the following statements about hyperparameters tuning with Gaussian Processes (GP) and Random Search (RS) are true?

- [x] RS can be faster than GP because RS is easier to parallelize.

Correct
You don't need to synchronize the parameters in RS.

- [ ] You should use GP if you have a lot of computational servers and every evaluation of the function is cheap/free.

Un-selected is correct

- [x] GP can be faster than RS because it suggests where to search for the next point using current information and uncertainty estimates.

Correct
GP can give you good results twice as fast as RS (in number of function calls), especially if the number of hyperparameters to optimize is relatively small.

---

1 / 1 point

7.
Bayesian optimization with Gaussian Process uses the following information about the optimized function $f(x)$.

- Only gradients of the function $\nabla f(x)$.
- Values of $f(x)$, gradients $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$.
- Values of $f(x)$ and gradients $\nabla f(x)$.
- Only values of $f(x)$.

Correct

---

1 / 1 point

8.
Which of the following problems you would use Bayesian Optimization for?

- [x] Find the best geographic coordinates for oil producing station

Correct
Cost of each probe is really high

- [ ] Find optimal weights in the Logistic Regression model.

Un-selected is correct

- [x] Find a molecule that possesses certain properties (drug discovery)

Correct
An application of Bayesian optimization to this problem was discussed in the last lecture.

- [x] Optimize the configuration of the neural network: number of neurons in each layer, parameters the optimization algorithm.

Correct
Bayesian optimization is a possible solution here.

Gaussian Processes and Bayesian Optimization

Quiz, 8 questions